MOLECULAR MODELING AND SIMULATION OF BACTERIAL CHEMOSENSORY ARRAYS

BY

C. KEITH CASSIDY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Yann R. Chemla, Chair
Professor Klaus Schulten, Director of Research
Professor Zaida Luthey-Schulten
Professor S. Lance Cooper
Professor John D. Stack

Abstract

The movement of an organism in response to environmental chemical cues is known as chemotaxis. Motile bacteria use chemotaxis to navigate through their environments, enabling cells to efficiently locate favorable growing conditions while avoiding harmful ones. Central to this ability, bacteria posses a universally conserved sensory apparatus, known as the chemosensory array, which involves the clustering of thousands of proteins into a highly cooperative signaling network. The present dissertation will present my work using techniques in computational modeling and simulation to investigate the molecular structure and function of the bacterial chemosensory array. A brief overview of each chapter follows.

Chapter 1 provides an introduction to the systems-level features of chemotaxis in the model organism *Escherichia coli* as well as an overview of the molecular organization and function of the chemosensory array.

Chapter 2 gives an outline of the core methodologies used in my work, specifically all-atom molecular dynamics (MD) simulation and Molecular Dynamics Flexible Fitting (MDFF). In addition, two of the primary techniques used to analyze the MD simulations presented in this dissertation are sketched out, namely structural clustering based on root-mean-square displacement (RMSD) [1] and Principal Component Analysis (PCA).

Chapter 3 reports my work, in collaboration with Peijun Zhang's Lab, to investigate the structural and dynamical features of the extended chemosensory array [2]. Using computational techniques to synthesize multi-scale structural data from X-ray crystallog-

raphy and cryo-electron tomography (cryo-ET) [3], an atomic model of the cytoplasmic portion of the chemosensory array from *Thermotoga maritima* is constructed and refined. Through the use of large-scale MD simulations [4], a novel conformational change in a key signaling protein is identified and subsequently shown to be critical for chemotaxis signaling in live *E. coli* cells.

Chapter 4 details the construction of an atomic model of a complete, transmembrane (TM) chemoreceptor. In particular, I use homology modeling and MD simulations, informed by biochemical and X-ray crystallographic data, to derive a model of the *E. coli* serine receptor (Tsr), including the previously uncharacterized TM four-helix bundle and HAMP domains. In addition, I report a series of MD simulations of a fragment of the resulting Tsr model, investigating the structural and dynamical effects of mutations on a key control cable residue. Preliminary MD simulations of the intact Tsr model are also presented.

Chapter 5 reports work in collaboration with Michael Eisenbach's Lab at the Weizmann Institute, exploring the role of acetylation on CheY activation and the generation of clockwise (CW) flagellar motor rotation [5]. Specifically, I present a series of MD simulations that investigate the effect of a hyperactivating mutation at a key acetylation site and offer a molecular explanation of acetylation-dependent generation of CW flagellar motor rotation.

I conclude with a brief description of recent work, expanding upon the results of the previous chapters, which has resulted in the first atomically resolved model of the *E. coli* transmembrane chemosensory array.

To Elisabeth,

for her constant light in the dark places.

Acknowledgments

In the happiest of circumstances, one is able to express gratitude directly to an advisor upon the pair's dissolution. Regrettably, I do not have this opportunity. Hence, I would like to here, first and foremost, extend into the ether my eternal thanks to Professor Klaus Schulten for his tireless generosity and guidance. Anything that I was able to accomplish in the pages that follow was made possible only though the professional opportunities and personal inspiration that Klaus provided. For these, I can offer compensation only through the extension of the work that he and I began and through the faithful transmission of his spirit of curiosity and reverence for nature to those around me. Though the water is gone, the well will not run dry.

I am also deeply grateful for the collaborators and colleagues with whom I have had the immense pleasure of working alongside and learning from over the course of my degree. These include especially Juan Perilla, Boon Chong Goh, Wei Han, Ben Himes, Oshri Afanzar, Davi Ortega, Ariane Briegel, Sandy Parkinson, and Peijun Zhang. For helping along an "outsider" such as myself, many in the bacterial chemotaxis community deserve thanks, including Nick Bartelli, Igor Zhulin, Rick Dahlquist, Brian Crane, Joe Falke, and Jerry Hazelbauer. I'd also like to thank all of the current and former members of the TCBG, in particular Jodi Hadden, Rafael Bernardi, Wen Ma, Chaitanya Sathe, Chris Chipot, David Hardy, Jim Phillips, and John Stone, for providing a world-class intellectual environment and technical infrastructure in which to work and play. For always getting me where I needed to go, I greatly appreciate the holders of the keys to the kingdom: Kelly Young, Donna Fackler, and Nancy Mallon.

I've managed also to gain—perhaps by hook, though probably by crook—a number of wonderful friends, which have rendered possible the existential survival of graduate school. For these services, my boundless best goes out to Ray Clay, Doug Packard, Juan Garcia, Juan Perilla, Glen Goodman, and Ryan Musa. I also tip my coolest hat to the baddest subset of patrons from Kopi, Quality, and Esquire (I sincerely hope I've made it plain in other ways who they are). Together they've brightened many a night.

To Elisabeth Reynolds, the singular interaction of this epoch, I express my profound indebtedness. Thank you, Liz, for the years we shared—for 707, for NOLA, for 507. Thank you for your patience and kindness, for helping me to personally instill these qualities within, and for teaching me how to love myself. Thank you with all that I have. I wish to thank also the entire Reynolds family for their unwavering support and for the opening of their homes to me. For bestowing upon me affection greater than any other organism on the planet, I heartily acknowledge The Hound and his poor judgment.

Finally, I would like to thank my family for their ever-present love and encouragement, for their sincere interest in my life, and for teaching me the value of hard work and the power of hope in things greater than ourselves. As I reflect on the immensity of the undeserved support and warmth I've received, I can't help but feel like the luckiest of so-and-so's.

This work was supported by the National Institutes of Health grant 9P41GM104601 as well as the National Science Foundation (NSF) grants PHY0822613 and ACI-1524703. This research is part of the Blue Waters sustained-petascale computing project, "The Computational Microscope" NSF PRAC award ACI-1440026. In addition, this research used the Stampede supercomputer at the Texas Advanced Computing Center (TACC), provided by grant MCA93S028 from the Extreme Science and Engineering Discovery Environment (XSEDE).

Contents

Chapter 1

Background

1.1 Bacterial Chemotaxis

The ability to sense, process, and appropriately respond to cues from the environment is essential to the survival of any organism. Chemotaxis is the name given to the fundamental and ubiquitous capacity of biological cells to translate environmental chemical stimuli into motile behavior [6]. Indeed, chemotactic mechanisms govern cell movement and migration in a wide variety of biological processes, including fertilization, differentiation, and immune response in eukaryotes [6–10] as well as foraging and pathogenesis in prokaryotes [6,11–15]. Motile bacteria, in particular, use chemotaxis to locate nutrients and avoid toxins by monitoring and integrating changes in the ambient concentrations of specific chemicals, so-called chemoeffectors, and using this information to affect the movements of motility-related appendages (e.g., flagella or cilia) [11,16,17].

In the model bacterium *Escherichia coli*, the organism in which chemotaxis is best understood, cells respond to their chemical environment in a binary fashion: either randomly change direction (tumble) or keep going (run). This behavior is described thoroughly and eloquently in Howard Berg's book on the subject [18]. Here I will summarize the key aspects. As depicted in Figure 1.1.A, in the absence of a chemoeffector gradient, *E. coli* cells stochastically wander through their environment in a random walk of interspersed runs and tumbles. In the presence of a chemoeffector gradient, this random walk becomes positively biased through the extension of runs in favorable directions (i.e.,

directions in which chemoattractant concentrations are increasing or chemorepellent concentrations are decreasing). In this way, cells are able to move appropriately (on average) along a chemoeffector gradient (Fig. 1.1.B). Remarkably, *E. coli* is able to simultaneously integrate many, potentially conflicting, chemical signals within its complex aqueous environment, enabling the cell to efficiently locate optimal living conditions. How, though, does *E. coli* decide whether to run or tumble and when to do so?

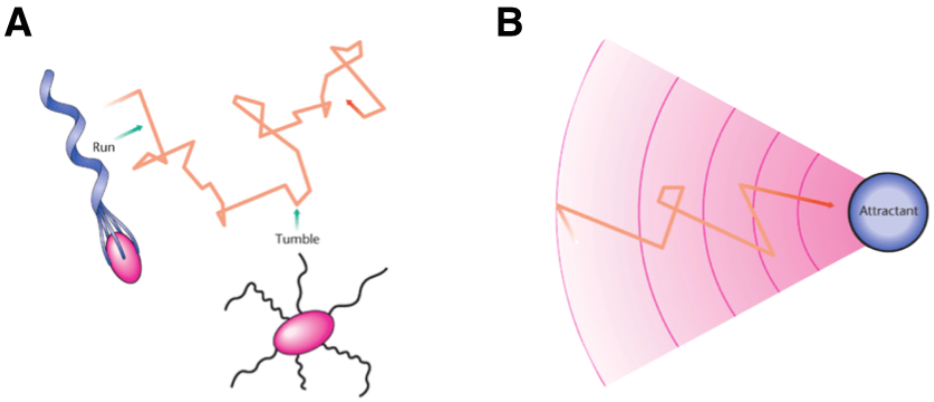

Figure 1.1: (A) In the absence of a chemoeffector gradient, *E. coli* cells undergo a random walk, either swimming smoothly in a straight line (Run) or erratically turning in a new direction (Tumble). (B) *E. coli* biases its motile behavior in response to chemoeffector gradients (shown here as a chemoattractant gradient). By extending the length of runs up the attractant gradient, the cell is able to move on average towards the attractant source. Figure modified with permission from [6]

1.2 The *E. coli* Chemotaxis Network

Sophisticated networks of proteins underlie the ability of bacteria to make sense of what they encounter in their environment and to convert this information into a decision regarding their movement. Though the details of these networks vary somewhat from species to species, the general chemotaxis control mechanism is conserved throughout bacterial diversity and involves a so-called two-component signal transduction system, the most common form of bacterial signal transduction [19, 20]. Typically, within a two component system, a sensory signal regulates the autophosphorylation activity of a ho-

modimeric histidine kinase, which in turn, transfers a phosphate group to a response regulator protein that triggers a corresponding cellular response [21]. The two-component chemotactic network of *E. coli*, pictured schematically in Figure 1.2, has provided the basis for the now decades-long investigation of bacterial chemotaxis and represents one of the most thoroughly studied behavioral systems in biology [6, 11, 16, 17, 22–24]. Here, I will briefly sketch out the key molecular players and their topological interactions.

The protein components of the *E. coli* chemotaxis network comprise and couple two intricate protein machines: the chemosensory array and the flagellar motor. Environmental chemical information is obtained by the network through specific protein receptors, termed chemoreceptors (red), on the surface of the cell's inner membrane. The binding of chemoeffector to a chemoreceptor leads to the transduction of signals across the cell membrane, which are used to regulate the autophosphorylation activity of a histidine kinase CheA (blue) with the help of an adaptor protein CheW (green). In particular, the binding of chemoattractant encourages inactive CheA, while an attractant-free chemoreceptor favors active CheA in which ATP is used to produce transferrable phosphate groups. The response regulator CheY (yellow) accepts a phosphate group (maroon) from activated CheA and diffuses through the cytoplasm to bind one of several

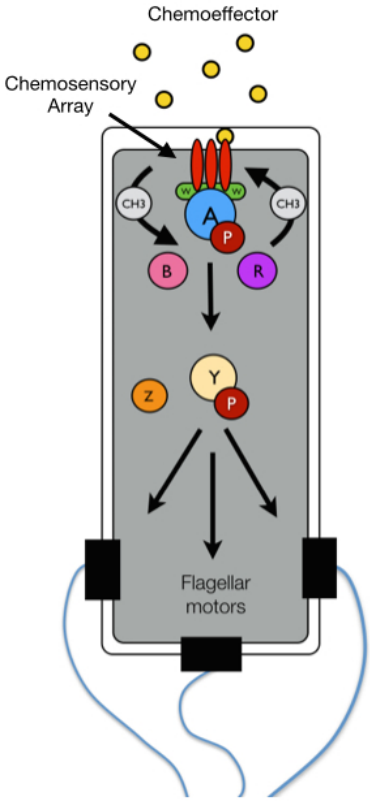

Figure 1.2: The *E. coli* Chemotaxis Network. Schematic of proteins involved in the coupling of environmental chemical gradients to cellular swimming pattern.

flagellar motors. The binding of phospho-CheY to a motor shifts its rotational bias from the default counter-clockwise (CCW) setting to clockwise (CW). CCW motor rotation gives rise to a tightly packed bundle of flagella, which favors runs, while CW motor rotation causes the flagellar bundle to break apart and induces tumbling (Fig. 1.1.A). Hence, through chemoreceptor-mediated kinase regulation, cells couple the physical binding of chemoeffector to the rotational bias of their flagellar motors and ultimately cellular swimming pattern.

In addition to the protein components described above, which are universal among chemotactic bacteria, several other proteins assist with the regulation of sensory signals in *E. coli* [25]. In particular, CheZ (orange) removes the phosphate group from phosphorylated CheY, effectively suppressing the chemical connection between CheA and the flagellar motors. Additionally, two other enzymes, CheR (purple) and CheB (pink) [26, 27], enable cells to tune their chemotactic sensitivity to stimulus intensity via the so-called adaptation process. Through the addition (by CheR) or removal (by CheB) of methyl groups at several specific amino acid sites on chemoreceptors, their ability to regulate CheA activity is drastically affected [28, 29]. Importantly, due to the relatively slow rate of receptor (de)methylation (compared to ligand binding), chemoreceptor methylation levels provide a record of the cell's immediate chemical past, giving rise to a kind of short-term molecular memory [30,31]. In this way, the cell compares its past environment with the one currently being reported by its chemoreceptors to decide whether things are getting better or worse.

1.3 The Chemosensory Array: Structure and Function

Central to their sensory function, chemoreceptors cluster along with CheA and CheW to form highly-ordered, supramolecular complexes known as chemosensory arrays [31,32].

Strikingly, the global architecture of the chemosensory array, schematized in Figure 1.3, appears to be a universally conserved feature in chemotactic bacteria and archaea, forming an extended honeycomb lattice $10^4 - 10^5$ nm^2 in size and containing more than 10^4 proteins [33–35]. Within this lattice, chemoreceptors (red circles) trimerize to form so-called trimers-of-dimers (TODs) [31]. Two TODs combine with a single CheA dimer (blue) and 2-4 CheW monomers (green) to form a core signaling unit (CSU, Fig. 1.3), the minimal physical unit necessary for proper CheA activation and regulation [36]. Utilizing highly specific interfaces between CheA and CheW, many CSUs then cluster to form an extended lattice of hexagonally packed TODs about rings consisting of (1) both CheA and CheW and (2) only CheW [2,37–39]. Each ring type forms the center of a distinct hexagonal organization (Figure 1.3), which we will refer to as the *CheA-trimer* and *CheA-hexamer* respectively (referring to the organization of CheA within the hexagon).

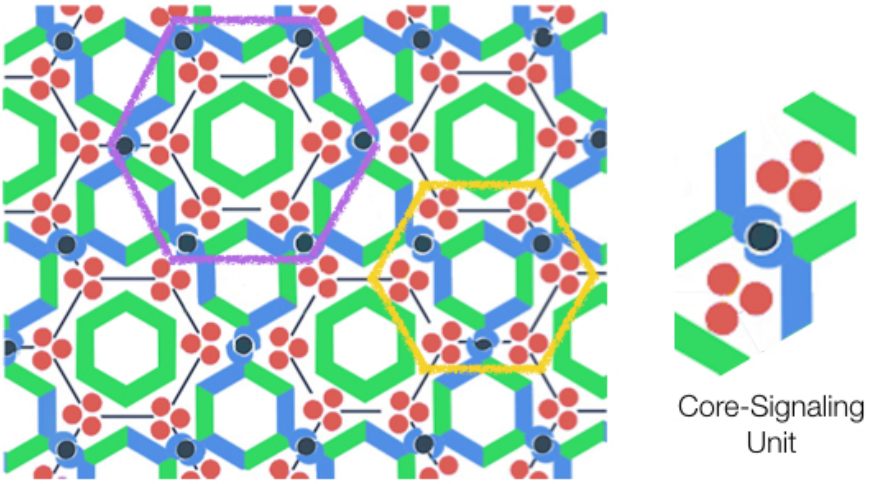

Figure 1.3: Schematic of extended chemosensory array architecture. An array trimer and hexamer are highlighted in yellow and purple respectively. A core-signaling unit is isolated to the side. Figure modified with permission from [38].

The extensive molecular interaction networks formed within the CSU and extended chemosensory architecture place the control of each CheA kinase under the joint regulation of many chemoreceptors, giving rise to a highly cooperative signaling response.

This cooperativity gives rise to a number of exceptional information processing features, including drastic signal amplification and precise adaptation [11, 40–42]. In particular, *E. coli* can amplify sensory signals up to 50-fold, allowing cells to sense extremely minute changes in concentration—less than three molecules per cell volume [41, 43, 44]. In addition, *E. coli* cells exhibit precise adaption (i.e., after some time, swimming pattern is accurately reset to its pre-stimulation default), permitting a sensitive response to chemical gradients over an incredibly wide range of background chemical concentrations—up to five orders of magnitude [40, 44, 45]. How do cells accomplish this remarkable computational feat? My work aims to elucidate, with atomistic detail, the mechanisms underlying the cooperative transduction and regulation of sensory signals within the *E. coli* chemosensory array, to better establish the link between the molecular and systems level features of bacterial chemotaxis. Importantly, the structural conservation of the extended array architecture over immense evolutionary distances suggests strongly a significant preservation of signaling mechanism and will, with hope, greatly extend the scope of the work presented in this dissertation.

Chapter 2

Methodology

2.1 Molecular Dynamics: Algorithm and Force Field

The principal method used in our study of the bacterial chemosensory array is molecular dynamics (MD). All-atom MD simulations provide an atomically resolved characterization of the structure and dynamics of biomolecules and their complexes. As such, it has become a powerful tool—indeed, a computational microscope—for studying the phenomena of molecular biology, including protein folding, drug-substrate interaction, and functionally-related conformational changes in biomolecules [4, 46].

The MD method works by computing the movements of atoms in a molecular system through the numerical solution of a series of coupled, second order differential equations, representing the physical and chemical aspects of the system [47]. The form of these equations, for each atom i, is given by Newton's Second Law,

$$m_i \frac{d^2 \overrightarrow{r}_i}{dt^2} \;=\; \overrightarrow{F}_i \;=\; -\nabla_{\overrightarrow{r}_i} U_{\mathrm{MD}}(\overrightarrow{R}) \quad ; \quad i = 1, 2, ..., N \tag{2.1}$$

where m_i and $\overrightarrow{r}_i$ are the mass and spatial coordinate vector of atom i respectively. The function $U_{\mathrm{MD}}(\overrightarrow{R})$, defined in Equation 2.2, describes the potential energy of the system and depends on the coordinates $\overrightarrow{R} = \{\overrightarrow{r}_1, \overrightarrow{r}_2, \ldots, \overrightarrow{r}_N\}$ of all N atoms. Through Equation 2.1, the acceleration of each atom can be determined by computing the total force $\overrightarrow{F}_i$ acting on it due to its interactions with the rest of the system (as specified by U_{MD}). The re-

sulting accelerations can then be used to update the positions and velocities of each atom according to a prescribed numerical integration scheme such as the Verlet algorithm [48]. In practice, the above process is repeated millions or even billions of times, giving rise to a collection of temporal snapshots known as an MD trajectory, which can be analyzed for interesting patterns.

$$\begin{aligned} U_{\mathrm{MD}}(\overrightarrow{R}) &= \sum_{bond} k_i^{bond}(\overrightarrow{r}_i - \overrightarrow{r}_0)^2 + \sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2 \\ &+ \sum_{dihedrals} k_i^{dihedral}[1 + cos(n_i\phi_i + \delta_i)] \\ &+ \sum_i \sum_{j>i} 4\varepsilon_{ij}[(\frac{\sigma_{ij}}{\overrightarrow{r}_{ij}})^{12} - (\frac{\sigma_{ij}}{\overrightarrow{r}_{ij}})^6] + \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 \overrightarrow{r}_{ij}} \end{aligned} \tag{2.2}$$

In classical, all-atom MD simulations, atoms are idealized as spheres of charge with given masses and atomic radii. The function $U_{\mathrm{MD}}(\overrightarrow{R})$, along with the set of parameters it contains, define the so-called MD force field, which mathematizes the interatomic potentials characterizing the physical and chemical interactions between the atoms comprising the system. The first three terms on the right hand side of Equation 2.2 describe 'bonded' interactions (i.e., interactions between atoms that are covalently attached to one another). As schematized in Figure 2.1, bonded interactions come in three varieties: *bonds*, *angles*, and *dihedrals*, which are typically treated as parameterized quadratic (bonds and angles) or sinusoidal (dihedrals) functions. The last two terms in Equation 2.2 describe 'nonbonded' interactions, in particular, the van der Waals (term 4) and long-range electrostatic (term 5) potentials. These are approximated via Coulomb's law and a parameterized Lennard-Jones 6-12 potential, respectively [48]. In general, the parameters for and between each atomic species are derived from experiments or higher-level quantum-mechanical simulations and represent (ideally) that species' average atomic and molecular behavior in a variety of chemical contexts.

The MD simulations presented in this dissertation utilize NAMD [49], a high performance MD software package, taking advantage of cutting-edge computing hardware and the latest petascale infrastructures to enable simulations of millions of atoms over microseconds and beyond [3]. In addition, the state-of-the-art molecular graphics package, VMD [50], was used for constructing the molecular models presented here as well as for the visualization and analysis of MD trajectories.

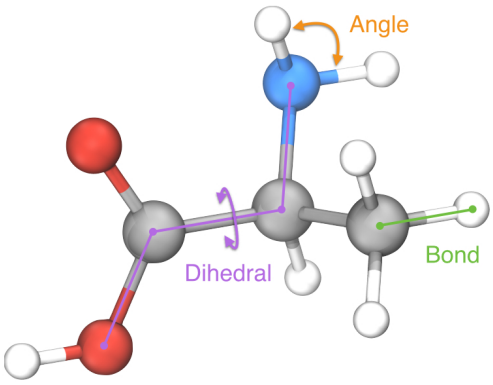

Figure 2.1: Bonded Interactions in MD Force Field. Alanine residue with examples of a *bond*, *angle*, and *dihedral* interaction labeled.

2.2 Molecular Dynamics Flexible Fitting

In addition to standard MD simulation, my work makes use of another computational method, namely Molecular Dynamics Flexible Fitting (MDFF) [51–53], to explicitly integrate all-atom MD simulations with multi-scale experimental structural data from X-ray crystallography, NMR, and electron microscopy (EM). As schematized in Figure 2.2, the effect of an MDFF simulation is to drive the conformation of a high resolution (1-4 Å) protein structure towards some other conformation represented in a low-to-intermediate resolution (10-40 Å) EM density. Importantly, the amenability of large protein complexes to *in vivo* and *in vitro* structural characterization by EM, albeit at lower resolution, permits the use of MDFF to construct from the resulting densities, atomically resolved models of molecular complexes in their native organizations—structures otherwise experimentally intractable [2–4, 54].

Briefly, the MDFF method works by deriving an external potential $V_{\mathrm{EM}}(\vec{r})$, to be applied to each atom, based on the relative strengths of electron density in different re-

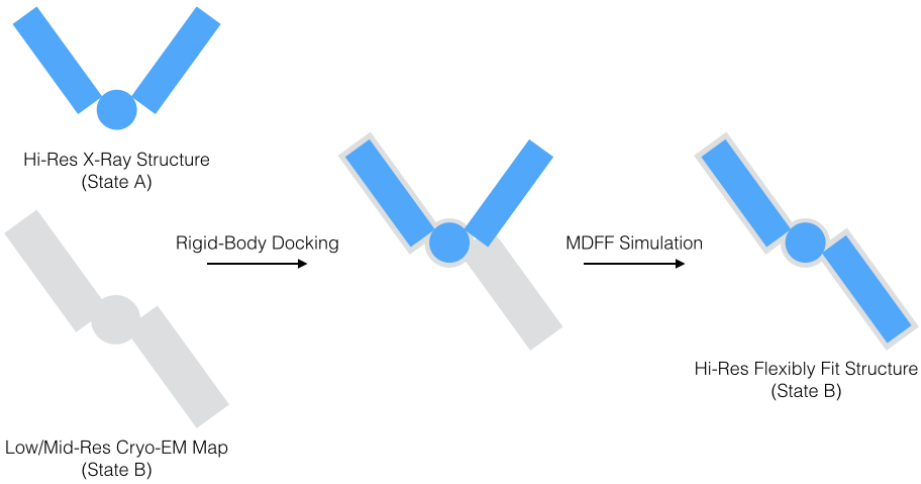

Figure 2.2: Schematic of MDFF Workflow. A high resolution X-ray or NMR structure in some conformation A is rigidly docked into an EM density of the same structure in another conformation B. MDFF enables the use of an MD simulation to computationally flexibly match the structure to the density, achieving optimal overlap between the two.

gions of the EM data. As shown in Equation 2.3, $V_{\mathrm{EM}}(\overrightarrow{r})$ is constructed from the *inverted* electron density $\Phi(\overrightarrow{r})$ such that regions of high density (representing well structured portions of a biomolecule) are more attractive, while regions of low density (representing disordered biomolecular regions) are neutral. To reduce noise, a threshold density value Φ_{thr} is selected below which the density is ignored. In addition, the overall all strength of the potential can be scaled using the β parameter. The resulting data-derived potential is then added to the MD force field, and this combined potential is used in Eq. 2.1 for the MD force calculation.

$$V_{\mathrm{EM}}(\overrightarrow{r}) = \begin{cases} \beta(1 - \dfrac{\Phi(\overrightarrow{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}) & \text{if } \Phi(\overrightarrow{r}) \geq \Phi_{thr} \\ \beta & \text{otherwise} \end{cases} \tag{2.3}$$

2.3 Conformational Clustering of MD Trajectories

This section is reproduced in part with permission from John E. Stone, Juan R. Perilla, C. Keith Cassidy, and Klaus Schulten. GPU-accelerated molecular dynamics clustering analysis with openACC. *Parallel Programming with openACC*, Elsevier, pp. 215-240, 2016.

Molecular dynamics (MD) simulations give rise to a collection of structural snapshots of a molecular system, known as a trajectory, that can be analyzed for interesting patterns [55, 56]. Due to the physical principles that govern MD simulations, an ensemble of molecular conformations near a local free-energy minimum is sampled and transitions are commonly observed between nearby low-energy basins (Fig. 2.3). Determination of the conformations at the minima of basins is important in the study of protein dynamics and for elucidating the connections between molecular structure and biological function [3, 5, 56, 57].

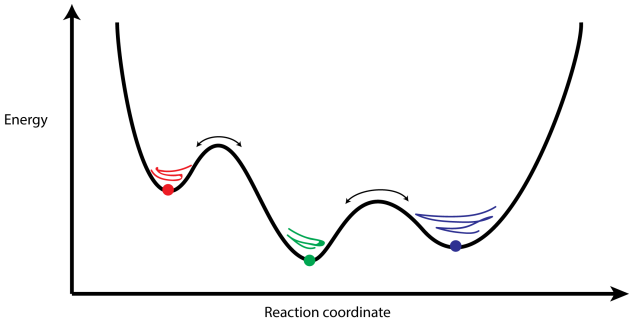

Figure 2.3: Schematic of underlying molecular free-energy landscape sampled by an MD simulation. Molecules transition back and forth, as signified by black arrows, between local low-energy basins along a reaction coordinate. Solid circles denote the representative conformations of each energetic basin. Figure reproduced with permission from [1].

The structural classification of an MD trajectory using clustering analysis provides a way to systematically determine the conformations associated with separate local free-energy basins. Clustering analysis is an unsupervised machine-learning technique that attempts to meaningfully group data based on the concept of proximity or similarity [58]. In the context of MD simulations, molecular conformations are characterized by an array of points in a three-dimensional space, with each point marking the location of an atom in the molecule. Hence, a common method for measuring the dissimilarity between

conformations is the root-mean-squared deviation (RMSD) given by

$$\mathrm{RMSD}_{i,j} = \frac{1}{S}\left[\sum_{l}^{S}\left(\vec{x}_i - \vec{x}_j\right)^2\right]^{\frac{1}{2}}, \tag{2.4}$$

where S is the number of selected atoms in the molecule. Though, RMSD is not a metric in the strict sense (e.g., it does not obey the triangular inequality), it nevertheless succinctly captures the similarity or dissimilarity between structural features of the molecules. In particular, conformations whose structural features are similar yield small RMSD values (< 1.5 Å), while dissimilar features are punished severely.

In general, clustering algorithms principally fall into one of two categories: either hierarchical or partitional, which differ primarily in the way in which clusters are determined [58]. In particular, hierarchical methods organize data into a hierarchical tree of nested clusters using either an agglomerative or divisive scheme [58]. Agglomerative schemes work in a bottom-up fashion, forming larger clusters from smaller ones, while divisive methods use a top-down approach. The decision of whether to merge or divide a given cluster is made according to a so-called linkage criterion that is specified by the user. Partitional methods, on the other hand, organize data into non-overlapping groups. Such methods require the user to specify the number of desired clusters, usually denote as k. Based on this parameter, an objective function particular to partitional method in use is iteratively optimized to arrive at a final grouping of k clusters [58].

Both hierarchical and partitional clustering employ a measure of pairwise proximity or similarity between elements in the input dataset. This information is commonly represented as a similarity (or dissimilarity, or "distance") matrix, in which the ijth-element of the matrix gives the value of the similarity measure between elements i and j in the data. In the case of MD simulations, a trajectory containing S selected atoms and N coordinate frames may be viewed as a 3×S×N time-evolving coordinate matrix. Utilizing

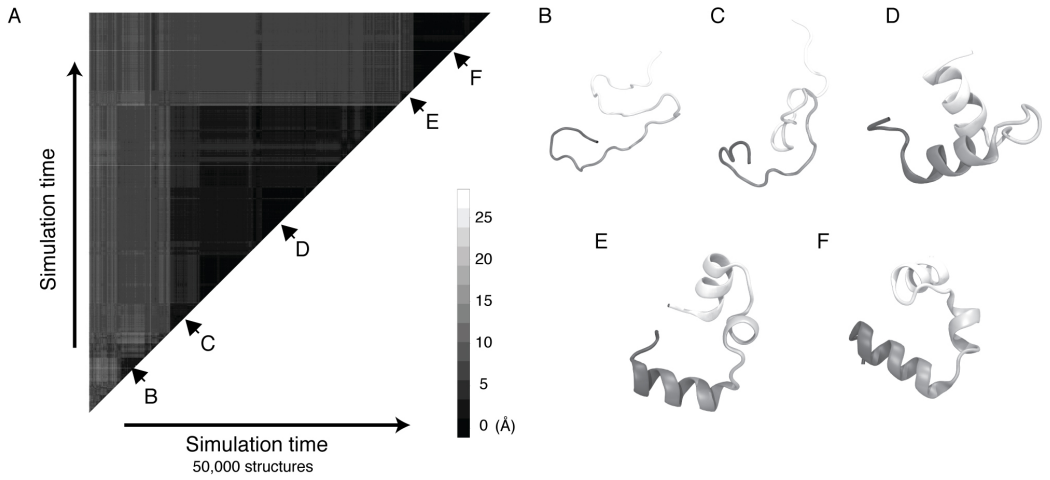

Figure 2.4: Computational folding of the villin headpiece protein from an initial random configuration [55, 59]. (A) Root-mean-squared deviation (RMSD) matrix between pairwise structures in an MD trajectory. Color-scale ranging from black to white, represents RMSDs between 0 to 25 Å. (B,C) Unfolded structures of the villin headpiece, because of the intrinsic flexibility of unfolded proteins the molecule is able to sample multiple extended states. (D,E) Intermediate states observed during the folding simulations. Structural motifs are present in the form of helices, however, the relative orientation of such helices is different from the canonical fold. (F) Folded structure of the villin headpiece reached after 5 μs of MD simulation [55]. Figure reproduced with permission from [1].

RMSD to characterize the structural similarity between molecular conformations within a trajectory (Eq. 2.4), an RMSD dissimilarity matrix can be constructed for use in clustering analysis. As depicted in Figure 2.4.A, the RMSD dissimilarity matrix may be represented as an upper-triangular matrix in which each element RMSD_{ij} gives the pairwise RMSD between structures $\vec{x}_i$ and $\vec{x}_j$. Importantly, during a simulation the molecule of interest diffuses freely, therefore translational degrees of freedom are removed from the resulting trajectories. Similarly, the structures must be aligned in order to remove any rotational degrees of freedom, this alignment is normally accomplished by finding the rotation matrix for which Eq. 2.4 is minimal [60, 61].

2.4 Principal Component Analysis

Principal Component Analysis (PCA) is a non-parametric, linear algebra technique that constructs a set of linearly uncorrelated variables (i.e., principal components, PCs) from observational data using an orthogonal transformation [62]. Importantly, the PCA procedure results in PCs that most efficiently describe the variance or "spread" of the original data. That is to say, the first PC maximally explains the variance of the data, while the second PC describes as much of the remaining variance as possible in a direction orthogonal to the first PC, and so on. As correlations between variables in complex, high-dimensional data can often render certain dimensions redundant, projection of the data onto a subset of the PCs can be used to effectively reduce the dimensionality of the data and expose simplified underlying patterns [63, 64].

Mathematically, PCA works as follows: First, correlations between each pair of variables in the data set are quantified using the *covariance* measure given by

$$cov(X, Y) = \frac{\sum_{i=1}^{T}(X_i - \bar{X})(Y_i - \bar{Y})}{T - 1}, \tag{2.5}$$

where $\bar{X}$ and $\bar{Y}$ are the mean values of variables X and Y respectively, and T is the total number of samples. In the context of MD, the variables X, Y, etc. represent atomic positions and T is the total number of frames (i.e., temporal snapshots) in a trajectory. The covariance between each pair of variables are then collected into the so-called covariance matrix given by

$$C = \begin{bmatrix} cov(X_1, X_1) & cov(X_1, Y_1) & cov(X_1, Z_1) & \dots & cov(X_1, Z_N) \\ cov(Y_1, X_1) & cov(Y_1, Y_1) & cov(Y_1, Z_1) & \dots & cov(Y_1, Z_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ cov(Z_N, X_1) & cov(Z_N, Y_1) & cov(Z_N, Z_1) & \dots & cov(Z_N, Z_N) \end{bmatrix},$$

where N is the total number of atoms considered and (X_n, Y_n, Z_n) are the (x, y, z) positions of atom n. Hence, C is a $3Nx3N$ square, symmetric matrix. Finally, we factorize C, using singular value decomposition, into a form represented by its eigenvectors and eigenvalues. For a real, symmetric matrix this decomposition takes the form

$$C = \Omega \Lambda \Omega^T, \tag{2.6}$$

where Λ is a *diagonal* matrix whose elements are the $3N$ eigenvalues and Ω is a square, *orthogonal* matrix whose ith column is the eigenvector corresponding to the ith eigenvalue. In the above scheme, the PCs are just the normalized eigenvectors of C and the corresponding eigenvalues specify the fractional variance (i.e., the percentage of the variance of the original data) explained by each PC. Hence, to form a low-dimensional representation of our data, we simply choose a subset of eigenvectors with the highest eigenvalues and project our data onto the space spanned by these vectors.

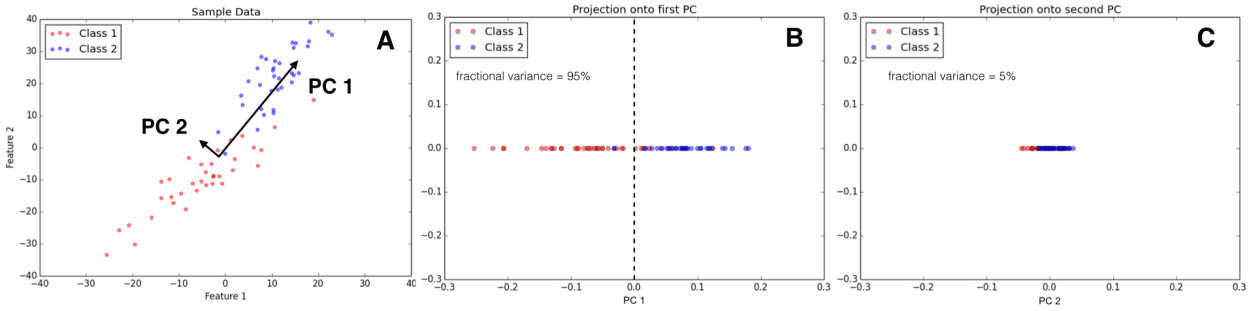

Figure 2.5: Example of Principal Component Analysis (PCA). (A) Sample data structured as two classes with strong positive correlation. The principal components (PCs) from PCA are shown. The length of the vector specifying a particular PC is proportional to the variance in the data described by that PC. (B) Projection of the original data onto PC 1. PCA respects the basic two class structure of the data as highlighted by the dashed line. (C) Projection of the original data onto PC 2.

As a visual example, consider the set of two-dimensional (2D), class-structured data shown in Figure 2.5.A. One clearly sees that the data is primarily spread along a line running 45° between the *feature 1* and *feature 2* axes. These are examples of "strongly cor-

related" variables, as the change in the value of one is closely related to the change in the value of the other (in this case an increase/decrease in *feature 1* is, in general, accompanied by an increase/decrease in *feature 2*). In addition we see that the two classes of data (colored either red or blue) are roughly separated with *class 1* being, on average, lower and to the left and *class 2* being higher and to the right.

PCA of this data set yields two, orthogonal PCs as drawn in Figure 2.5.A. The projection of the original data onto the first and second PCs are shown in Figure 2.5.B and Figure 2.5.C respectively. As expected, PC 1 accounts for much of the original variance (with a fractional variance of 95%), such that a one-dimensional representation retains much of the separation between individual data points, while PC 2 explains a much smaller amount (in this case, the remaining 5% since our original data was 2D). In addition, we see that PCA is sensitive to the class structure of the data, with the projection onto PC 1 neatly separating the majorities of *class 1* and *class 2* to the left and right of the origin respectively. Hence, one could reasonably replace the original 2D, (*feature 1*, *feature 2*) representation by a single number, namely the projection onto PC 1 and still retain the majority of the information in the data set. In very high dimensional data such as MD trajectories, however, one cannot visualize such correlations and PCA becomes a powerful exploratory tool.

Chapter 3

The *T. maritima* Chemosensory Array: Modeling and Simulations of Cytoplasmic Portion

Reproduced in part with permission from C. Keith Cassidy, Benjamin A. Himes, Frances J. Alvarez, Jun Ma, Gongpu Zhao, Juan R. Perilla, Klaus Schulten, and Peijun Zhang. Cryo-EM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. *eLife*, 4:e08419, 2015.

3.1 Introduction

Chemotactic responses in bacteria require the assembly of large, membrane-bound complexes of sensory proteins, known as chemosensory arrays, to mediate the signal transduction that ultimately controls cell motility (see Chapter 1 for more detail). The fundamental physical unit of chemosensory signaling, the so-called core signaling unit (CSU, Fig. 3.1.A), is comprised of six chemoreceptor homodimers, organized as two trimers-of-dimers (TOD), a single CheA homodimer, and 2-4 CheW monomers [36]. The hierarchical organizations of the CSU and extended array architecture (Fig. 3.1.B) give rise to a highly cooperative signal processing apparatus, which endows the basic two-component chemotaxis infrastructure with heightened information processing and control capabilities [30, 32, 42, 43, 65].

As depicted in Figure 3.1.A, chemoattractant is sensed by binding to the periplasmic domains of chemoreceptors (red), initiating sensory signals that are transduced across the cell's inner membrane and over more than 30 nanometers to the binding sites of CheA

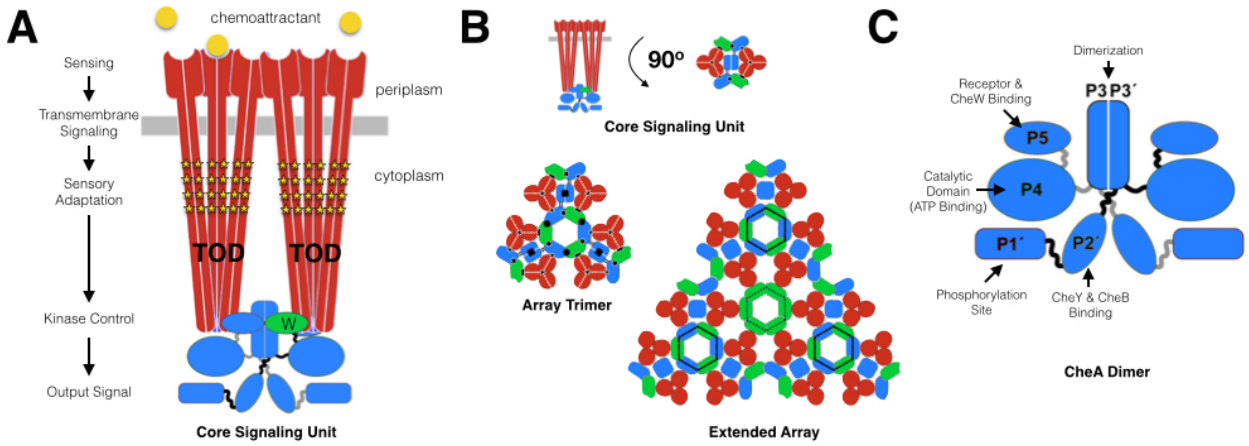

Figure 3.1: Overview of functional organization within chemosensory array. Chemoreceptors, CheA, and CheW are shown in red, blue, and green respectively. (A) The core signaling unit (CSU) integrates sensory signals from six chemoreceptors, organized as two trimers-of-dimers (TOD), to control the autophosphorylation activity of the CheA kinase. Signals are regulated via the reversible methylation of specific residues (stars) in the adaptation region of chemoreceptors. (B) Many CSUs cluster using highly specific interactions between CheA and CheW to create a large allosteric network that further integrates sensory signals. (C) The histidine kinase CheA functions as a homodimer, with each protomer containing five, distinct domains (P1-P5) whose functions are listed. Figure modified with permission from [31].

(blue) and CheW (green) on the receptor's cytoplasmic tip. Signals from three chemoreceptors are combined through inter-receptor interactions within TODs, two of which are coupled across the CheA kinase. Through the reversible methylation of specific glutamyl residues (stars) in the sensory adaptation region of their cytoplasmic domains, chemoreceptors regulate signals arising from chemoattractant occupancy, effectively deriving a composite signal that is used to appropriately regulate (through the help of CheW) the autophosphorylation activity of the CheA kinase. Remarkably, in addition to receptor-mediated kinase regulation within a single CSU, the aggregation of many CSUs to form the extended array architecture (Fig. 3.1.B) creates an allosteric network in which a single receptor can influence the activity of as many as 35 CheA enzymes [66, 67].

To enable the transduction of sensory information downstream to the flagellar motors, the CheA histidine kinase autophosphorylates (i.e., transfers a phosphate group to) a self-contained HIS residue in response to sensory signals [68]. Each protomer of the CheA

homodimer (Fig. 3.1.C) consists of five domains (P1-P5), with each domain performing a distinct role in CheA enzymatic function. In particular, the P1 domain contains the substrate HIS residue to which a phosphate group (from hydrolyzed ATP) is first passed in a trans (i.e., inter-subunit) fashion. The P2 domain binds CheY and CheB to encourage the transfer of phosphate from P1 to the response regulators. The P3 domain provides the dimerization interface, and the P4 domain binds ATP and catalyzes the hydrolysis reaction. Finally, the P5 domain binds chemoreceptors and CheW, physically coupling these proteins and giving rise to the interfaces that allow the CSU and extended array to form (Fig. 3.1.A&B).

Important progress has been made in the structural characterization of the CSU and extended array, using a battery of genetic, biochemical, and biophysical techniques. In particular, atomic structures for key parts of the individual core signaling components (i.e., CheA, CheW, and chemoreceptors) [69–73] and several of their sub-complexes [38,71, 74] have been derived, the vast majority of these coming from the thermophilic bacterium *Thermotoga maritima*. In addition, a number of important interactions between the core signaling components have been deduced from soluble, multi-protein complexes [75–77] and in reconstituted, attractant-regulated core complexes [36, 78–81]. Nevertheless, the structurally localized and non-native nature of these data have made them difficult reconcile into a consistent structural picture.

Meanwhile, a global view of the extended array structure has emerged from cryo-electron tomography (cryoET) studies of native bacterial cells [33–35,37–39,82,83], establishing the overarching organization of the array as a universally conserved feature of bacterial chemotaxis (Fig. 3.2). Taken together, these studies have established the hexagonal lattice of TODs about kinase-filled and kinase-empty rings (Fig. 1.3) described in Chapter 1. Due to the thickness of cells, however, as well as cellular crowding and heterogeneity effects, past cryo-ET studies have been limited to discerning only the overall

arrangement of the core signaling components. Hence, despite the aformentioned strides toward a molecular description of the chemosensory array, a high-resolution structure of the intact, extended array has remained elusive, hindering a mechanistic understanding of the molecular events underlying sensory signal transduction and regulation.

In this chapter, I describe my work, in collaboration with Peijun Zhang's Lab at the University of Pittsburgh, to construct and refine an atomic model of the extended array structure. In particular, we have developed a novel reconstitution method yielding ultra-thin monolayer samples of the cytoplasmic portion of chemosensory arrays for visualization using cryoET. From these we have derived a three-dimensional density map of the reconstituted core signaling complex at 11.3 Å resolution using sub-tomogram classification and averaging. Next, through the computational synthesis of existing X-ray crystallography data and our new cryoET data, we have constructed an atomic model of the extended chemosensory array. Our model highlights novel interaction interfaces between chemoreceptors, CheA, and CheW and permits the use of large-scale, all-atom molecular dynamics (MD) simulations to further illuminate the molecular details of a key kinase-signaling event.

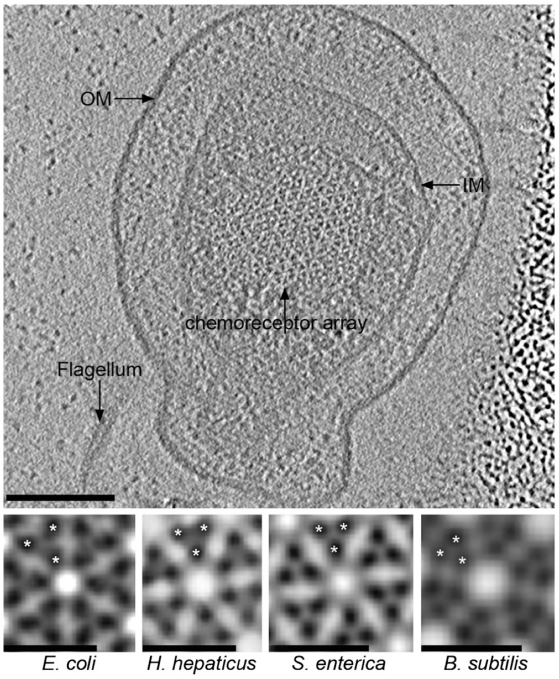

Figure 3.2: Extended array architecture as seen by cryo-electron tomography. (Top) Tomographic slice through the top of a *S. enterica* mini-cell. (Bottom) Subtomogram averages of arrays from several, distantly related species, revealing the universally conserved two-facing-two hexagonal organization of chemoreceptors. Figure reproduced with permission from [38].

3.2 Results

3.2.1 Reconstitution and cryoET of *E. coli* chemosensory array

Note: The experiments and analysis giving rise to the results presented in this section were carried out by members of Peijun Zhang's Lab (see author contributions in [2]). I have included them here in some detail as they are central for framing the computational work presented in subsequent sections and for better understanding the scientific significance of the work as a whole.

To overcome the limitations imposed by cellular tomography of native chemosensory arrays [33, 38, 39, 82, 83], we established an *in vitro* reconstituted system for high-resolution structural analysis of the signaling complex. Inspired by the template-directed method to assemble functional signaling complexes on lipid vesicles [84,85], we designed a Ni^{2+}-NTA lipid containing monolayer system (Fig. 3.3) [86, 87] to reconstitute the two-dimensional arrays of signaling complexes for structural analysis. We then expressed and purified to high homogeneity *E. coli* chemotaxis proteins: CheA, CheW, and a His-tagged cytoplasmic signaling domain of the wild-type (wt) Tar receptor (TarCF). His-tagged TarCF can be readily incorporated into the Ni^{2+}-NTA lipid monolayers. Through the addition of CheA and CheW, hexagonal lattices resembling the arrays of native cells were formed.

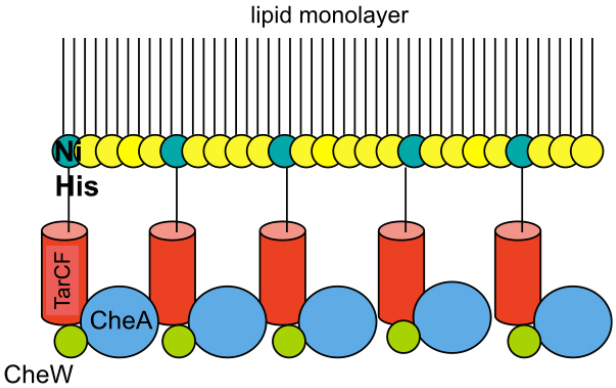

Figure 3.3: *In vitro* reconstitution of *E. coli* chemosensory array. Schematic of Ni^{2+}-NTA lipid monolayer system with CheA, CheW, and His-tagged TarCF, allowing for high-resolution structural analysis of the signaling complex by cryoET.

Compared to previous cellular tomography studies, the reconstituted monolayer system is ideal for high resolution structural analysis of chemosensory arrays by cryoET for several reasons: 1) the *in vitro* reconstituted monolayer array is thin (25 nm) and pseudo-crystalline, compared to cells with thicknesses ranging from 500 nm to 1 mm; 2) the monolayer arrays are reconstituted with purified components, hence the system is well-defined, in contrast to native arrays in the crowded cellular environment; 3) the reconstituted system allows for control over which array components are present as well as manipulation of their signaling state; 4) the in vitro system provides large numbers of sub-tomogram volumes ($\sim$3000 core signaling units/tomogram), thereby improving the noise statistics of the sub-tomogram averaging process central to achieving a high resolution structure.

Using cryoET, we collected and reconstructed, correcting for the contrast transfer function (CTF) of the microscope [88], 20 tomograms of monolayers containing reconstituted chemosensory arrays. Figure 3.4.A shows a typical raw tomographic slice (without CTF correction) of a reconstituted monolayer, illustrating patches of 2D lattices with information extending beyond 22 Å (inset, arrow). By extracting and classifying CTF-corrected sub-tomograms, centered on each hexagon of receptor TODs, we obtained two major classes of the receptor hexagons: one containing a trimer of core signaling units (Fig. 3.4.B and Fig. 3.5, cyan boxes) and one containing a hexamer of core signaling units (Fig. 3.4.C and Fig. 3.5, orange box), which we will subsequently refer to as the *CheA-trimer* and *CheA-hexamer*. Sub-tomograms within each of these classes were then averaged together, resulting in 3D density maps of the CheA-trimer (Fig. 3.4.B) and CheA-hexamer (Fig. 3.4.C) at 11.3 Å and 17.5 Å resolution respectively, as measured by gold-standard Fourier shell correlation (FSC) [89]. Notably, the best previously published tomographic data was 25 Å resolution [83]. In addition, by mapping the individual sub-tomograms from the above two classes onto the original contributing tomograms, we were able to

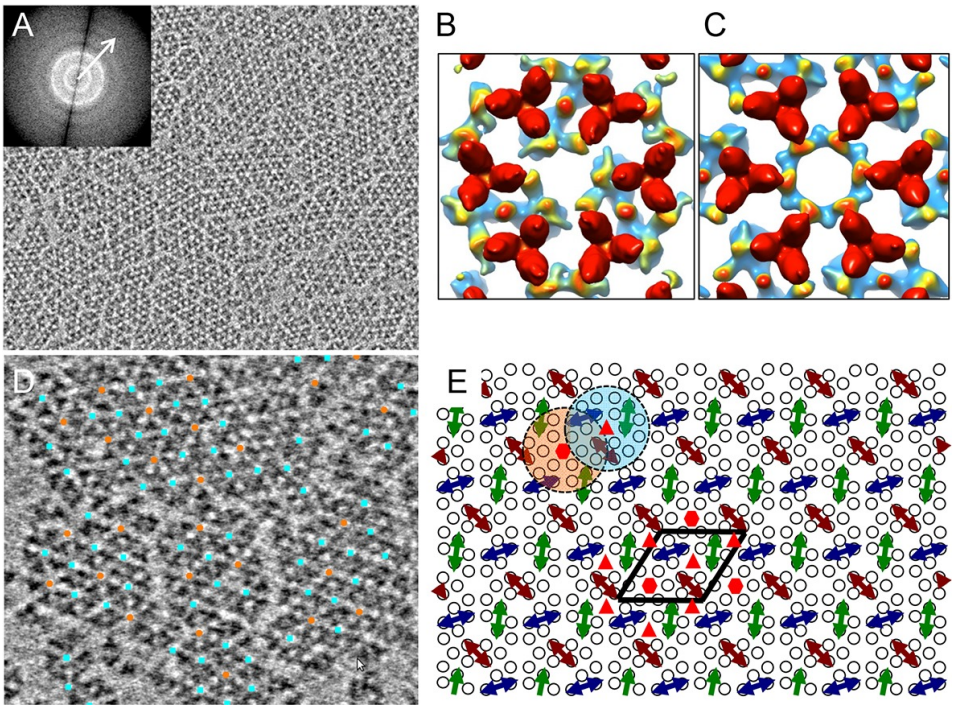

Figure 3.4: CryoET of monolayer arrays of TarCF/CheA/CheW ternary signaling complex. (A) A tomographic slice (1.2 nm thick) through the reconstituted monolayer arrays of TarCF/CheA/CheW. Inset, The Fourier transform of a selected region, displaying Thon rings with information extended to at least 22 Å resolution (arrow). (B&C) Averaged density maps of two sub-volume classes containing receptor hexagons (6 TODs) (red), one with a trimer of CheA dimers (CheA-trimer) (B) and the other with a hexamer of CheA dimers (CheA-hexamer) (C). Maps were generated following sub-tomogram volume classification and class-averaging, and are colored according to the height, from the receptor at the top (red) to CheA (blue) below. (D) Spatial arrangement of the CheA-trimer (cyan) and CheA-hexamer (orange) in the monolayer lattice array, after mapping the classified sub-volumes back onto the tomogram. The array is formed by interlocking CheA-trimer and CheA-hexamer subunits. (E) A schematic lattice model for the chemosensory arrays. Small circles represent receptor dimers; arrows represent CheA dimers (CheA2). Dashed cyan and orange circles highlight a CheA-trimer and CheA-hexamer respectively. The lattice unit cell is outlined in black.

extract the extended lattice organization of the subunits in the monolayer (Fig. 3.4.D), revealing an interlocking of the CheA-trimer and CheA-hexamer classes (Fig. 3.4.E) consistent with that seen in cellular tomograms.

3.2.2 Atomic model of *T. maritima* chemosensory array

The resolution of our cryoET data permitted the unambiguous assignment of distinct regions of density to specific protein components, enabling the construction of all-atom

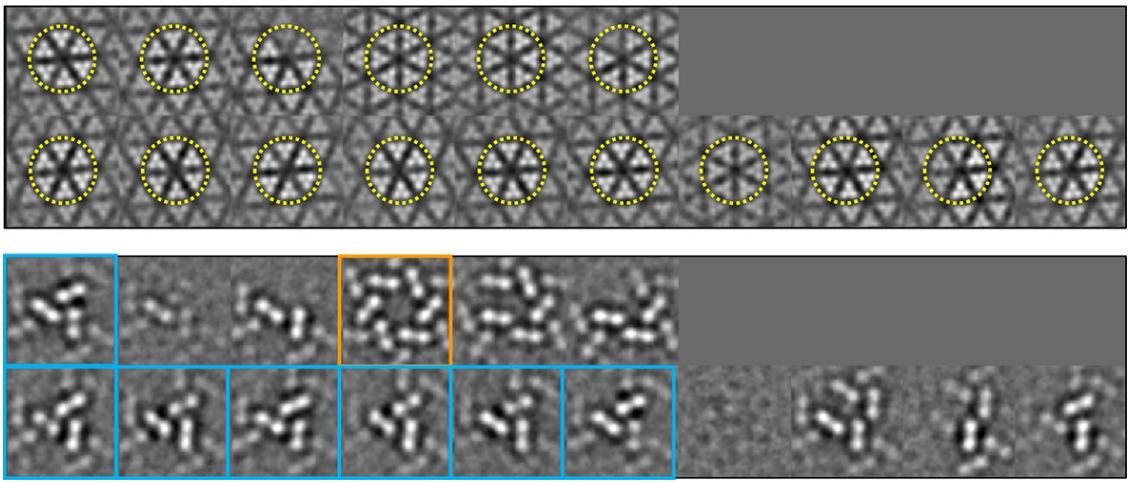

Figure 3.5: The sub-tomograms containing the receptor hexagon (6 TODs, yellow circles) were subjected to alignment and classification. Sections of the sub-tomogram classes are shown at the receptor region (top row in each set) and the CheA region (bottom row in each set). Two major configurations emerged, one with a trimer of CheA dimers (CheA-trimer, cyan boxes) and the other with a hexamer of CheA dimers (CheA-hexamer, orange box).

models of the chemosensory array substructures and extended lattice. An overview of the modeling and simulation procedures carried out in this study is provided in Figure 3.15. Briefly, we first constructed models of the chemoreceptor TOD, CheA-P3P4 dimer, CheA-P5/CheW ring, and CheW-only ring, taking advantage of existing high-resolution X-ray structures from the thermophilic bacterium *T. maritima* (Fig. 3.6). In particular, atomic coordinates of the cytoplasmic portion of the *T. maritima* receptor dimer were taken from the X-ray crystal structure of the TM1143 chemoreceptor (PDB 2CH7) [71]. Using as a reference homologous trimer-forming contacts from the X-ray structure of the cytoplasmic fragment of the *E. coli* Tsr TOD (PDB 1QU7) [69], a *T. maritima* receptor TOD model was obtained by appropriately aligning the individual TM1143 homodimer models. We then constructed an atomic model of the soluble *T. maritima* CheA dimer, including the dimerization (P3) and kinase (P4) domains based on atomic coordinates from the X-ray crystal structure PDB 1B3Q [70]. Finally, atomic models for both the CheA-P5/CheW and CheW-only rings were based on the X-ray crystal structure of the Receptor/CheA-P5/CheW ternary complex, PDB 4JPB [74]. In the case of the CheW ring model, the P5 domains of

the CheA-P5/CheW ring model were exchanged with CheW monomers, using the dual-SH3-like fold shared between by CheA-P5 and CheW, to obtain an appropriate placement and orientation with respect to the neighboring monomers. The TOD, CheA-P5/CheW, and CheW ring core component models were subjected to 150 ns of equilibration to ensure their structural integrity.

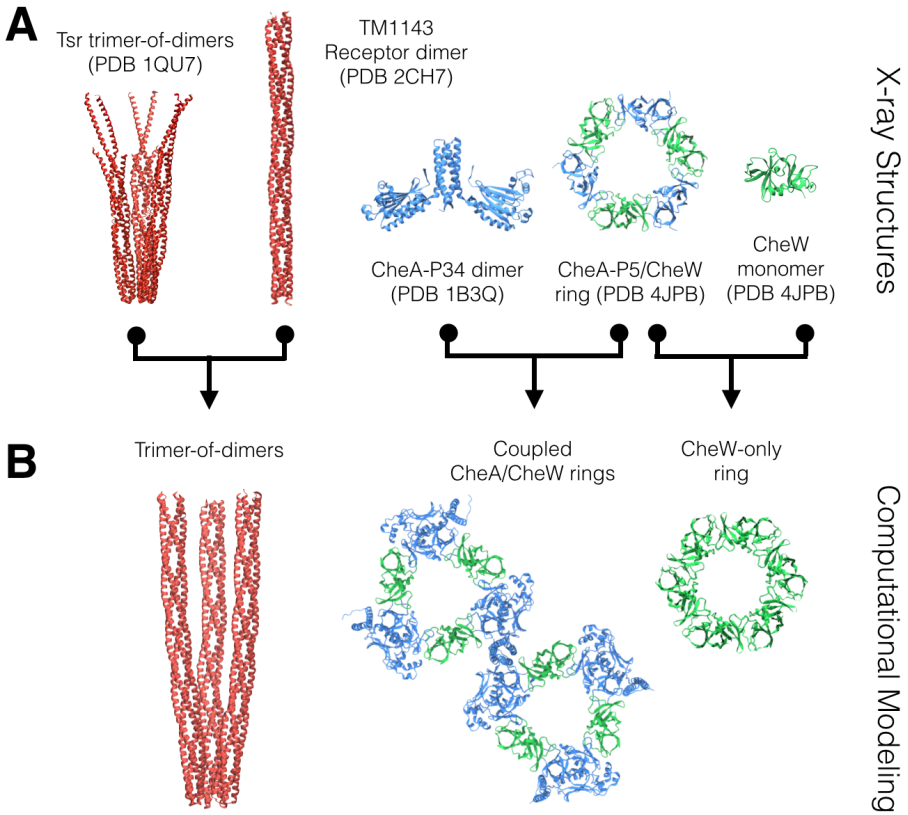

Figure 3.6: Atomic structures of core signaling components from *T. maritima*. (A) High resolution X-ray structures from *T. maritima* were taken as inputs for the generation of models corresponding to the array's core components. (B) Core components, namely the receptor trimer-of-dimers, coupled CheA/CheW rings, and CheW-only ring. The Protein Data Bank accession code for each template structure is given.

Models of the CheA-trimer (Fig. 3.7.A) and CheA-hexamer (Fig. 3.7.B) array substructures identified by sub-tomogram classification were then produced by heuristically-arranging the above core signaling component models in accordance with the extended protein organization evident in our density maps (Fig. 3.4.B&C) and assuming a 12 nm

lattice constant [33, 35]. Making use of the CheA-P5/receptor interface from the ternary complex structure PDB 4JPB [74], we modeled the CheW/receptor interface, assuming a receptor-binding mode homologous to that of CheA-P5. Using the CheA-P5 and CheW monomer/receptor models from the previous step, positional constraints on the receptor TODs were set relative to the height and orientation of the protein rings. Finally, CheA-P3,4 core component models were placed between adjacent TODs in accordance with the patterns observed in our density maps and joined to nearby ring-bound regulatory domains (P5) at the P4-P5 flexible linker.

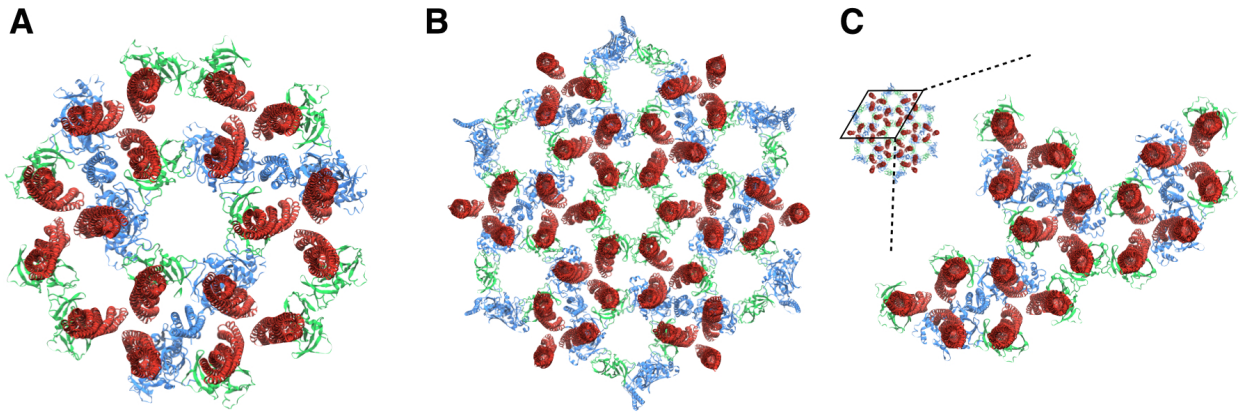

Figure 3.7: The *T. maritima* core component models were arranged heuristically, assuming a 12 nm lattice constant, to produce models of the array substructures: (A) CheA-trimer, (B) CheA-hexamer, and (C) unit cell. The CheA-trimer and array unit cell models were subsequently refined with all-atom MDFF and unbiased MD simulations respectively.

To further refine the protein conformations and inter-protein interfaces within our atomic models, we adopted a dual MD-based strategy, utilizing both unbiased MD and electron-density-biased molecular dynamics flexible fitting (MDFF) simulations [51–53]. For the subject of our unbiased refinement simulations, we extracted from the CheA-hexamer model a portion corresponding to the array unit cell, including six receptor TODs, three CheA dimers, and 12 CheW monomers arranged as three adjacently, coupled core signaling units (Fig. 3.16.C). The unit cell organization is particularly attractive as periodic boundary conditions can be used within our MD simulations to mimic the

bulk array (Fig. 3.7.C), preventing the need to interpret potentially problematic effects due to unconstrained boundaries. For our density-biased refinement simulations, we used both our 11.3 Å CheA-trimer and 17.5 Å CheA-hexamer density maps to computationally bias the tertiary structure of the protein components within those respective models. Solvation and ionization of the unit cell, CheA-trimer, and CheA-hexamer models produced systems of size 1.25 million, 1.75 million, and 4.58 million atoms, respectively, which were subsequently energy minimized and equilibrated for 10 ns, as described in the Methods section. The unit-cell model was then subjected to an 80 ns unconstrained production simulation, while the CheA-trimer and CheA-hexamer models were subjected to a 70 ns and 15 ns symmetry-constrained MDFF simulation, respectively. The resulting array substructures models agreed well with previous structural studies, in particular with respect to the residues participating in the CheA-P5/receptor and CheW/receptor interaction interfaces, as defined by NMR [76, 77, 90], crystallography [74], and disulfide mapping studies [79, 80]. For succinctness, specific residues participating in the various inter-protein interfaces within the array have been listed in Table 3.1 in the Supplemental Information section of this chapter.

Taken together our density maps and array substructure models also reveal a number of new features. First, with respect to the chemoreceptor TOD, our maps thoroughly resolved the individual receptor dimers (Fig. 3.8.A-C), further guiding the construction of our atomic TOD model. Indeed, although the cytoplasmic domain of the *T. maritima* TM1143 homodimer (Fig. 3.6) crystallized as a hedgerow [71], our model and simulations demonstrated that *T. maritima* receptors form a stable trimer-of-dimers (TOD) that maintain the conserved trimer-forming contacts observed in the *E. coli* Tsr TOD [69]. Moreover, our simulations revealed two additional salt bridges, namely E387/R389 (conserved as E402/R404 in *E. coli* Tsr) and E351/ R403 (structurally homologous to D363/R415 in *E. coli* Tsr), which further stabilize the TOD (Fig. 3.16). In addition, our density maps substan-

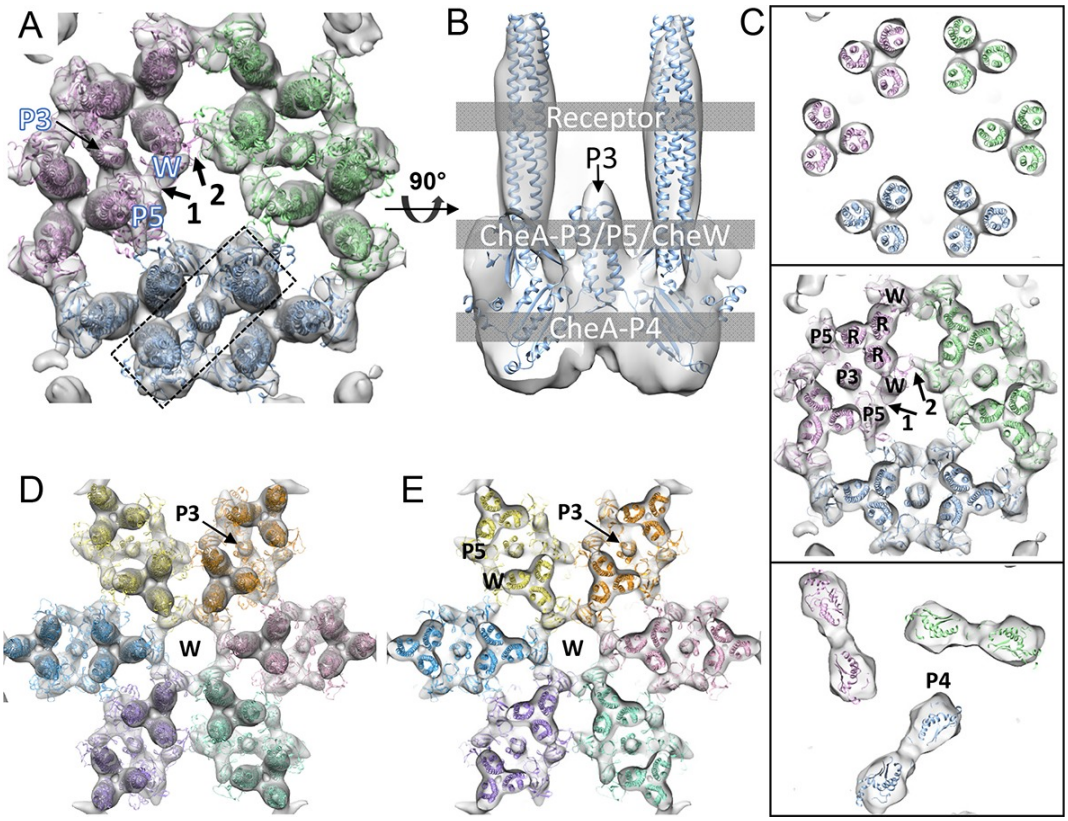

Figure 3.8: CheA-trimer and CheA-hexamer density maps with molecular dynamics flexible fitting (MDFF) of computationally constructed *T. maritima* subunit models. (A) Overall fitting of the CheA-trimer density map contoured at 1.5σ. The three core signaling units are colored in pink, blue and green. (B) A sectional view of the boxed region in A, rotated 90°. The protein components are labeled at the indicated height of the complex (gray boxes). (C) Sectional views of the gray-boxed regions in B at the receptor level (top), the CheA-P3 and P5/CheW ring region (middle), and CheA-P4 region (bottom). (D) Overall fitting of the CheA-hexamer density map contoured at 1.5σ. (E) A sectional view at the CheA-P3 and CheW-ring region of the CheA2-hexamer density map. In (A-E), CheA-P3, P4, P5, CheW and receptor are labeled as P3, P4, P5, W and R, respectively, and the CheA-P5/CheW interfaces 1 and 2 are indicated.

tially refine the locations of the individual domains of CheA within the context of the rest of the core signaling unit (CSU) (Fig. 3.8, Fig. 3.9.A). In particular, our maps clearly position the previously unobserved four-helix bundle of the CheA-P3 dimerization domain, showing that it runs parallel to the receptor and is positioned close to CheW-interacting receptor dimers (Fig. 3.8.A-C, Fig. 3.9.A). In tandem with this finding, our atomically resolved models revealed previously uncharacterized specific interactions between the P3 bundle and adjacent receptors, involving D333/K390 and D345/R379 contact pairs

(I304/N405 and D316/R394 in *E. coli* respectively) (Fig. 3.9.D). Furthermore, our maps dramatically refine the area of density projecting below the CheA-P5 domain, suggesting that the CheA-P4 kinase domain alone occupies this density region (Fig. 3.8.B&C, Fig. 3.9.A). Indeed, our MDFF simulations directed the CheA-P4 kinase domain to this region of density (Fig. 3.9.B), significantly enhancing the goodness of fit of the CheA dimer as a whole (Fig. 3.9.C). Finally, our densities show that the previously described 'empty hexagon' that is surrounded by six CheA-occupied hexagons [38] is not empty, but rather contains a well-ordered continuous ring of densities (Fig. 3.4.C). Our model and simulations showed that this density could be nicely explained by a ring of CheW (Fig. 3.8.D&E), a possibility previously speculated [39] to strengthen the interlocking CheA/CheW baseplate and contribute to the ultra-stability and high cooperativity of the array [91, 92]. An atomic model of our MDFF-refined CSU was deposited in the Protein Data Bank [93] under accession code 3JA6 (Fig. 3.14).

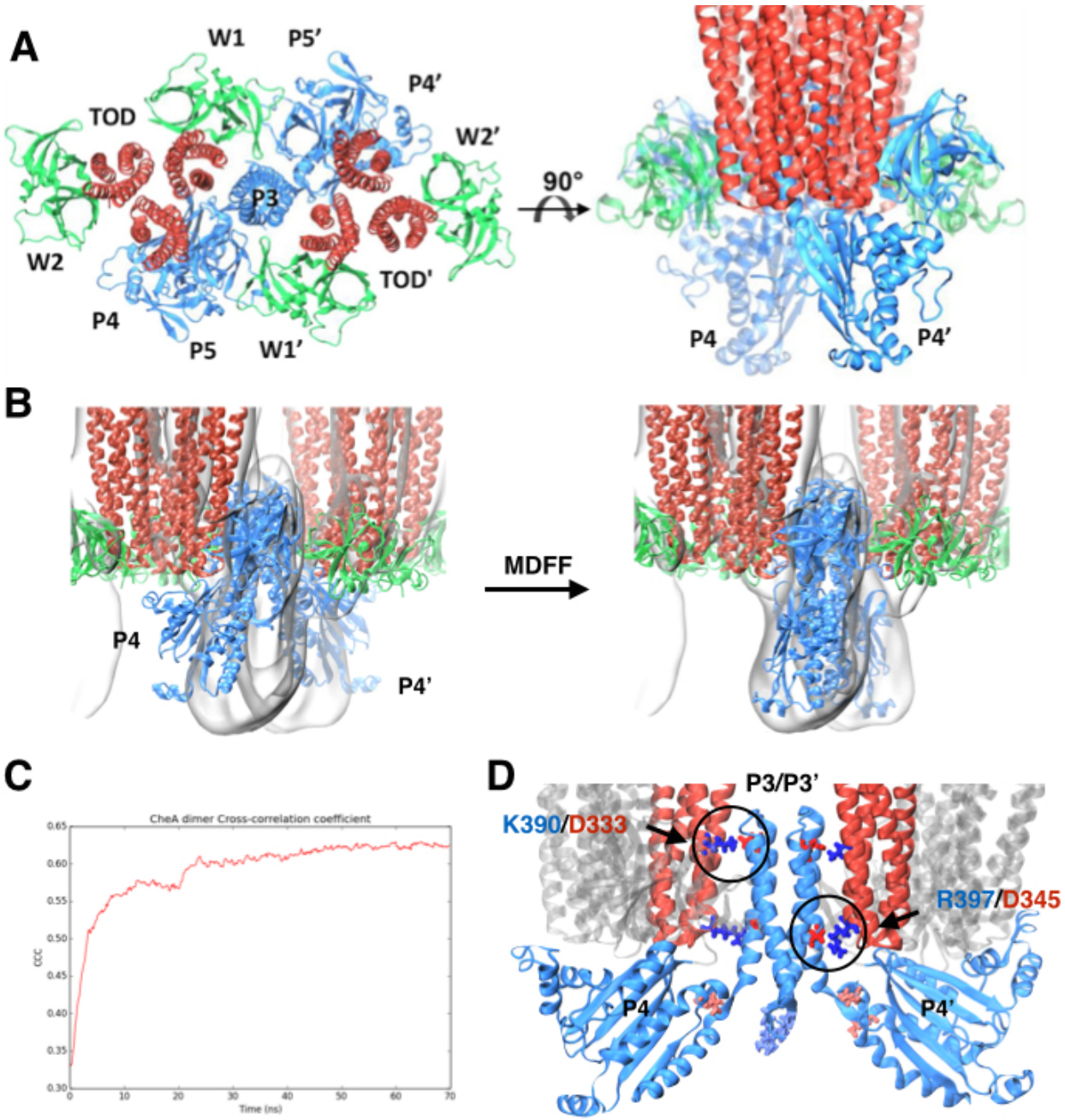

Figure 3.9: MDFF refinement of atomistic array models reveal new protein-protein interfaces. (A) 80 ns MDFF simulation of CheA-trimer refines inter-domain organization of CheA dimer, especially in key kinase (P4) region, altering its interactions with nearby receptors and CheW. (B) Cross-correlation coefficient between a CheA dimer and local region of CheA-trimer density map plotted as a function of time. The goodness of fit of the CheA dimer significantly increases over the course of the MDFF refinement simulation. (C) Atomic array models exhibits novel strong contacts (K390/D333 and R397/D345) between the CheA-P3 domain and neighboring receptor dimers that were previously unknown. CheA-P5, CheW, and non-interacting receptors have been grayed for clarity.

3.2.3 Identification of key conformational change in CheA kinase

The construction and refinement of atomic models of the array substructures permitted the use of equilibrium all-atom MD simulations to further investigate the molecular details of dynamic events potentially relevant to signaling. An overview of the key MD simulations conducted in this study is given in Figure 3.10. In particular, we ran extended, large-scale MD simulations on two systems: (1) the equilibrated unit-cell model (1.25 million atoms) and (2) the MDFF-refined CheA-trimer model (1.75 million atoms). We chose to focus our attention on the CheA-trimer model as the higher-resolution of its associated density map produced better resolved MDFF biasing forces and, consequently, a superior refinement over the CheA-hexamer. We then conducted a series of nine, 450 ns simulations of the unit-cell model and ten, 200 ns simulations of the MDFF-refined CheA2-trimer system for comparison.

Model	System	Simulation	Atom # (M)	Length (ns)
Core Component	Receptor TOD	Equilibration	0.55	150
	CheA-P5/CheW ring	Equilibration	0.13	150
	CheW ring	Equilibration	0.13	150
Array Subunits	CheA2-trimer	MDFF	1.75	70
	CheA2-trimer	Production	1.75	10 * 250
	CheA2-hexamer	MDFF	4.58	15
Unit Cell	Wild type	Production	1.25	80
	Wild type	Production	1.25	9 * 450
	R297A mutant	Production	1.25	9 * 450

Figure 3.10: Overview of all-atom molecular dynamics simulations conducted in this study with atom number reported roughly in millions (M) of atoms (including protein, solvent, and ions) and simulation duration in nanoseconds (ns). Where multiple, independent simulations were run, the duration is multiplied by the number of simulations.

Intriguingly, our simulations of both models revealed within various CSUs (Fig. 3.9.A) an ensemble of distinct conformations, including structures in which the associated CheA dimer displayed either an undipped conformation (Fig. 3.11.A) or dipped conformation

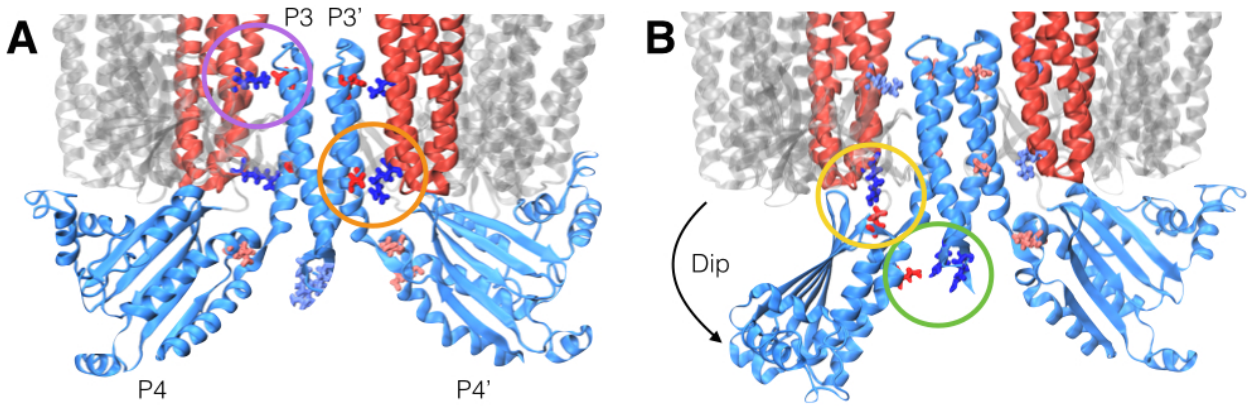

Figure 3.11: Representative CheA conformations from the (A) "undipped" and (B) "dipped" structural classes obtained from RMSD-based structural clustering. Strong, specific residue pairs stabilizing either conformation are indicated with colored circles. Specifically, K390/D333 (purple circle) and R397/D345 (orange circle) between the CheA-P3 domain and neighboring receptor were seen to stabilize the "undipped" conformation, while R297/E397 (green circle) between CheA-P3 and CheA-P4 as well as E390/R379 (yellow circle) between CheA-P4 and a nearby receptor tip were observed to stabilize the "dipped" conformation. CheA-P5, CheW, and non-interacting receptors have been grayed for clarity.

(Fig. 3.9.B). In the latter case, the P4 domain of one CheA monomer adopted a "dipped" state through rotations about the P3-P4 and P4-P5 flexible linkers, significantly affecting its contacts with neighboring receptor dimers and the P5 domain. As many biochemical, biophysical, and mutational studies have implicated dynamic structural changes within these regions of the core-signaling unit during the propagation of signals [79, 80, 83, 94], we systematically identified the distinct structural classes of CSU conformations present in our MD simulations and isolated them for comparative analysis. Specifically, we used the UPGMC hierarchical clustering method [63, 95] to assign the conformations of the 27 CSUs sampled in our unit cell simulations (3 CSUs/unit cell) to groups of similar structure based on their pairwise root-mean-square deviation (RMSD). An overview of the clustering process is given in Figure 3.17 in the Supplemental Information section of this chapter. Cross-examination of structures within the resulting core-signaling unit clusters revealed the formation of two new salt bridges stabilizing the "dipped" state, namely R297/E397 (R265/E368 in *E. coli*) between the P3 and P4 domains and E390/R379 (E361/R394 in *E. coli*) between the P4 domain and nearby receptor tip (Fig. 3.9.B). More-

over, to accommodate the reorientation of the P4 domain, the P3 dimerization bundle was observed to break the receptor contacts (D333/K390 and D345/R379) observed in the "undipped' state" (Fig. 3.9.A), suggesting that the mobility of the P3 bundle plays a key role in the conformational dynamics of the CheA dimer. A summary of key interactions occurring uniquely within the "undipped" and "dipped" conformations is given in Table 3.2 in the Supplemental Information section of this chapter.

We next sought to examine the temporal evolution of the dipping motion in each of the CheA dimers present in our simulations. For this purpose, we used Principal Component Analysis (PCA) [63] to systematically derive a pseudo reaction coordinate by which to easily monitor the CheA-P4 conformation. A total of four "dipping" events were observed in our unit cell simulations, as illustrated by projection of the conformations of the 27 CheA dimer time series onto the first principal component (Fig. 3.12.A). Importantly, an additional two dipping events were observed in the 30 CheA dimers of the relatively shorter simulations of CheA2-trimer model (Fig. 3.12.B), demonstrating that the ability of the conformational change to occur is not an artifact of the particular choice of CheA P4 positioning during modeling. Interestingly, the three extended "dipping" events observed in the unit-cell simulations (Fig. 3.12.A; red, blue, and green traces) as well as the two events observed in the CheA2-trimer simulations were accompanied by the formation of the R297/E397 contact. Notably, this contact was not formed in the one short dipping event, which returned to the "undipped" bulk state (Fig. 3.12.A; gold trace), suggesting that the R297/E397 contact may play a role in stabilizing the "dipped" state. To further investigate the significance of the R297/E397 contact for the conformational dynamics of CheA, we conducted nine additional, 450 ns unit cell simulations with an R297A mutation to prevent the potential formation of the R297/E397 salt bridge. Indeed, while two CheA dimers exhibited the dipping motion in these simulations, including one dimer that underwent two dips, the mutants quickly return to the bulk (Fig. 3.12.C).

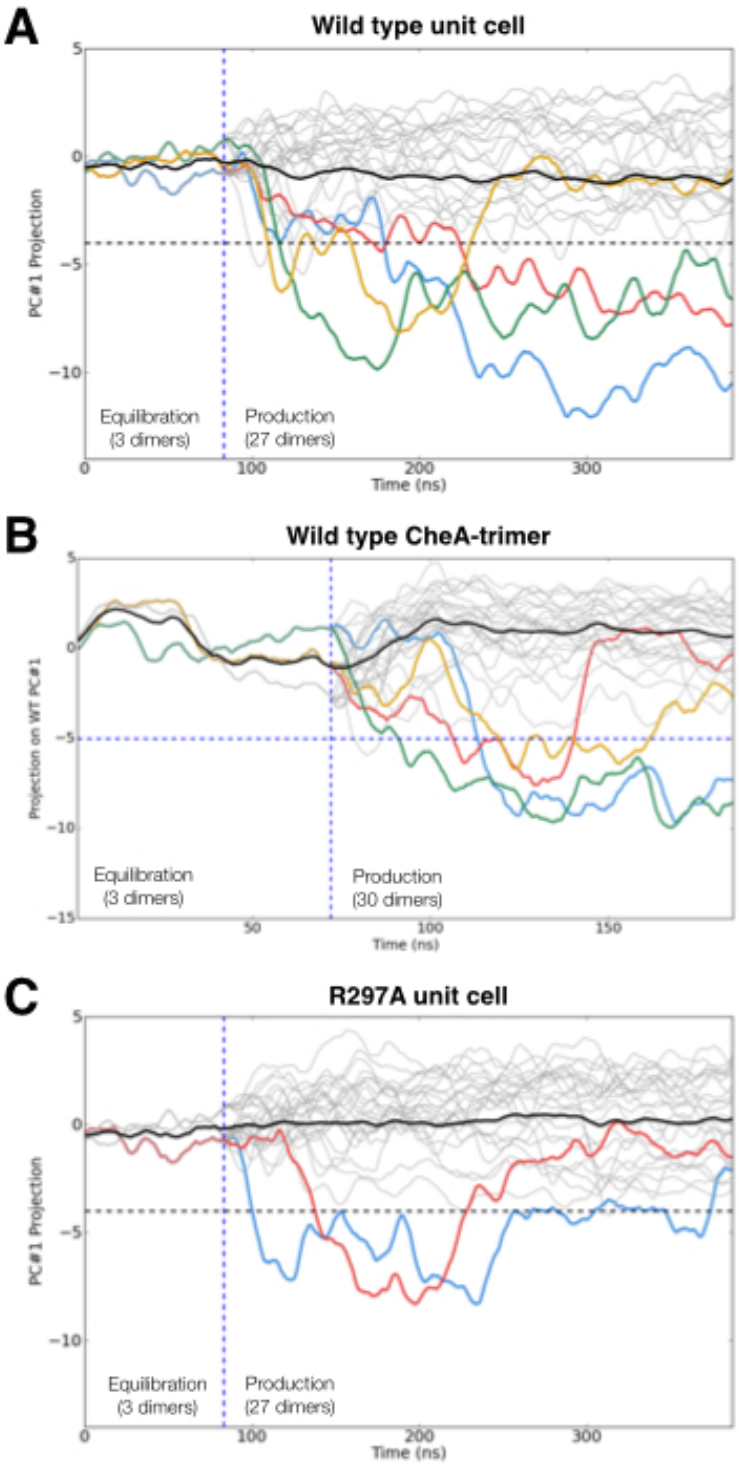

Figure 3.12: Time series of CheA dimer conformations extracted from simulations of (A) wild type unit cell, (B) CheA2-trimer, and (C) R297A mutant unit cell. Traces track the projection of the conformations of 30 CheA dimers onto the first principal component of the "dipping" motion. Colored traces track CheA dimers that undergo an extended (¿10 ns) "dipping" motion. Horizontal dashed line visually demarcates the "undipped" and "dipped" CheA dimer classes. Vertical dashed line separates equilibration and production simulations.

(Note: The experiments giving rise to the biochemical results presented below were carried out by members of Peijun Zhang's Lab, see author contributions in [2].)

To determine if the CheA-P4 dipping motion observed in the MD simulations of the *T. maritima* chemosensory array is sampled in the native chemotactic response of *E. coli*, we carried out cysteine disulfide cross-linking experiments. In particular, we tested the interaction interface for contacts existing in the "undipped" state (I304/N405 and D316/R394) or only in the "dipped" state (E361/R394) (Figure 5B). Notably, in the simulations, R394 of Tsr switches its contact with D316 of CheA-P3 to E361 of CheA-P4 during the transition of the CheA dimer from "undipped" to "dipped". Using soft-agar assays, it was seen that the chemotactic ability of the I304C/N405C double cysteine mutant is appreciably compromised compared to that of the control (cysless CheA/wt Tsr), any of the single mutants (I304C/wt Tsr, cysless CheA/N405C, cysless CheA/N405S), and when one half of the pair has been mutated to serine (I304C/N405S) (Fig. 3.13.A), suggesting that dynamic interaction between CheA-P3 and the receptor is important for chemotactic function. Moreover, *in vivo* cross-linking and western blot analysis showed a high molecular weight band present only in the double cysteine mutant, suggesting the presence of species formed by cross-linking between CheA-P3 and Tsr (Fig. 3.13.B). We also examined cross-linking residue pairs that involve Tsr-R394 interactions with CheA, one in the "undipped" state (CheA-E316C/Tsr-R394) and the other in the "dipped" state (CheA-E361C/Tsr-R394). When Tsr-R394 is replaced by a cysteine or serine, either as a single mutant (cysless CheA/R394C, cysless CheA/R394S) or in the context of a double mutant (D316C/R394C, D316C/R394S, E361C/R394C, E361C/R394S), the chemotaxis function of *E. coli* is partially inhibited. On the other hand, the chemotactic ability of CheA-E361C as a single mutant (E361C/wt Tsr) is also partially inhibited, while CheA-D316C (E316C/wt Tsr) mutation bears no effect on the function (Figure 5A). Furthermore, the cross-linking pattern of both Tsr-R394 mutant pairs showed two high molecular weight bands corre-

sponding to distinct cross-linked species, one with a disulfide formed between Tsr and CheA (Figure 5B, upper band, blue arrows) and the other with a disulfide between two Tsr molecules with the R394C mutation (lower band). Interestingly, the cross-linking of CheA-P4/Tsr (E361C/R394C) in the predicted "dipped" conformation is much weaker than the cross-linking of CheA-P3/Tsr (D316C/R394C) in the "undipped" state, though both involve the same R394 residue of Tsr. The lower cross-linking efficiency could be due to the relatively infrequent occurrence of the CheA "dipped" conformation, and/or because the residues are further apart in a dominant conformation, suggesting that the CheA-P4 "dipped" conformation observed *in silico* may have been sampled within the native chemosensory complex of *E. coli*.

Our MD simulations of the *T. maritima* unit cell further indicated that R297 on the CheA-P3 domain is potentially involved in the stabilization of the conformational transition of the CheA-P4 (Fig. 3.12.A&C). Indeed, substitution of the corresponding residue in *E. coli* (R265) with several amino acids of different properties (R265C/S/A/E) were all detrimental to the chemotactic function of *E. coli* as measured by the soft-agar assay, without affecting the cluster formation (Fig. 3.13.C). Since this residue is located at the N-terminus of the four-helix P3 dimerization motif, R265 could direct the P2-P3 linker away from the cis subunit and toward the trans subunit, thus anchoring CheA-P1,P2 to CheA-P4' for trans-interaction and phosphorylation [70]. A more complete model of the core-signaling complex for *E. coli* may be necessary to fully interpret the drastic impact of this single CheA residue on the entire chemotactic machinery.

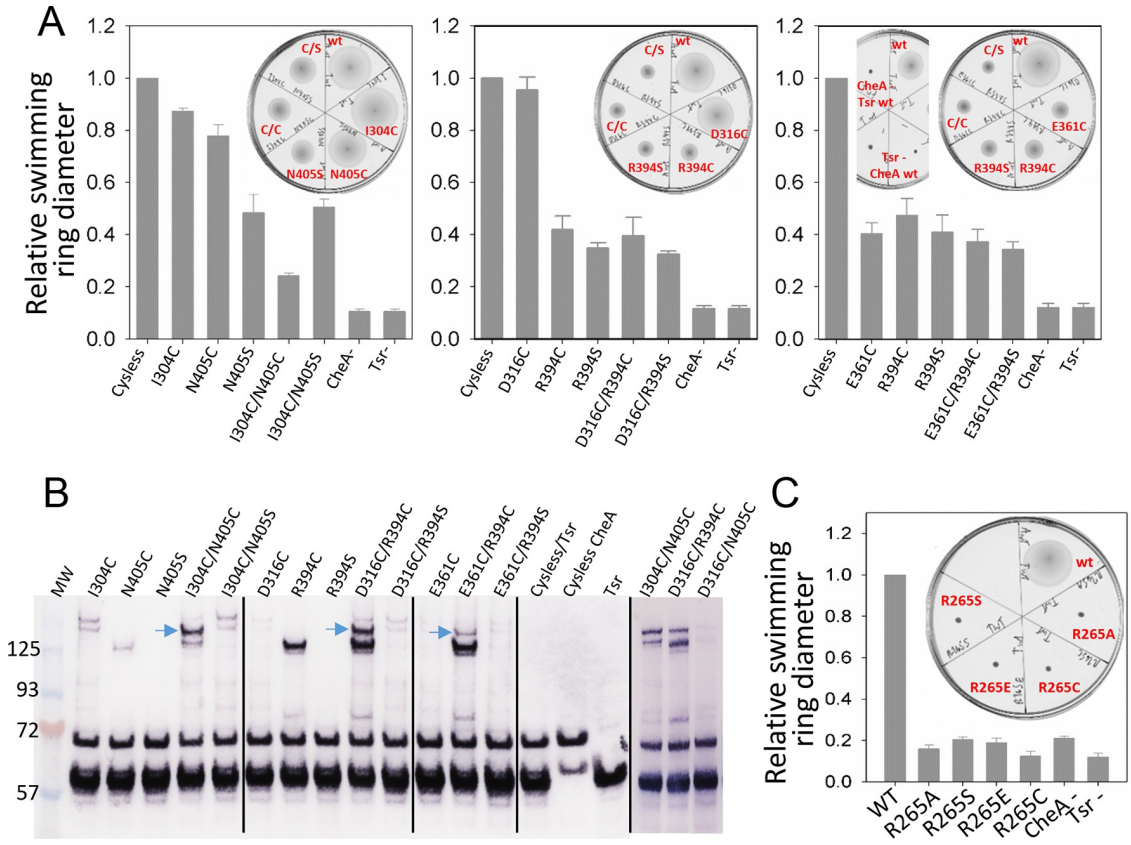

Figure 3.13: Biochemical validation of alternative CheA conformations in *E. coli* cells. (A) Swimming ability of *E. coli* cells with mutations in the CheA-P3 and Tsr interface (I304/N405 and D316/R394) and in the "dipped" CheA-P4 and Tsr interface (E361/R394). Swimming activities are normalized to the cysless CheA and wt Tsr, ± standard deviation (n=6). Inset, representative images of soft agar plates for swimming ability, with specific constructs labeled in red. (B) Disulphide cross-linking of the CheA-P3 and Tsr interface (I304C/N405C and D316C/R394C) in the "undipped" CheA dimer conformation (top panel of Figure 5B) and the CheA-P4 and Tsr interface (E361C/R394C) occurring in the CheA-P4 "dipped" conformation (bottom panel of Figure 5B). Non-reducing (top) and reducing (bottom) SDS-PAGE gels were analyzed by immunoblotting for Tsr and CheA. Cross-linked species were indicated with blue arrows. (C) Swimming ability of *E. coli* cells with mutations at R265 of CheA-P3 domain, normalized to the wt, ± standard deviation (n=8).

3.3 Discussion

In this study, we have utilized of a novel *in vitro* reconstitution technique to create ultra-thin monolayer samples of the chemosensory array, enabling the use of cryo-electron tomography to determine the most highly resolved three-dimensional structure of the *Escherichia coli* array to date. We were then able to use the electron tomography data to guide and refine the construction of the first atomically resolved model of the chemosensory array, using high resolution structures from *Thermotoga maritima*. Next, using large-scale, all-atom molecular dynamics, we identified a dipping motion of the CheA P4 domain, which was functionally characterized using swim assay and cross-linking experiments. While the role of the predicted conformational change in CheA is not immediately clarified in the preliminary biochemical experiments carried out here, our model highlights the importance of CheA dynamics for signaling and suggests that the dynamics of the P4 kinase domain, in particular, warrants special investigation. More importantly, the atomic model presented here, in general, provides improved knowledge of the positioning of the P3 and P4 domains, incorporates the presence of the CheW only ring, and identifies probable novel side-chain contacts within the extended chemosensory architecture. An atomic model of our MDFF-refined CSU was deposited in the Protein Data Bank under accession code 3JA6 (Fig. 3.14).

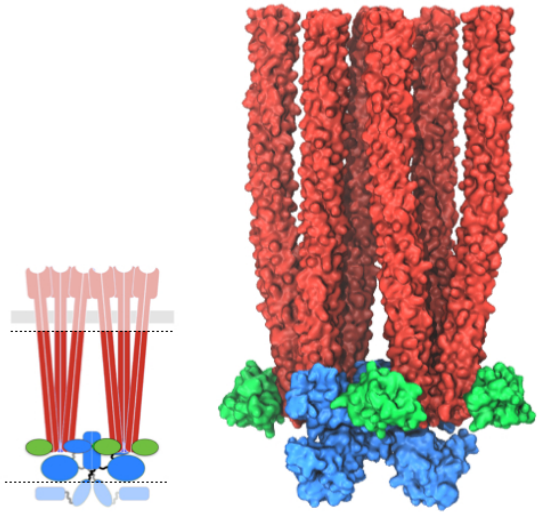

Figure 3.14: Model of core signaling unit (CSU). Atomic coordinates of a CSU from our MDFF-refined *T. maritima* CheA-trimer model were deposited in the Protein Data Bank under acession code 3JA6. A schematic of the portion of the native CSU covered by our model is situated to the left.

3.4 Methods

3.4.1 Molecular dynamics simulations

The array unit cell model was hydrated with TIP3P water molecules using VMD's solvate plugin [50], producing a simulation box defined by hexagonal lattice parameters a=208 Å, b=208 Å, c=334 Å, α=90°, β=90°, γ=120°. Using VMD's autoionize plugin, the hydrated system was then neutralized and subsequently ionized with sodium and chloride ions to the physiological concentration of 150 mM, resulting in a model containing 1,153,756 atoms. The unit cell model was then subjected to a series of conjugant gradient energy minimizations (300,000 steps in total) and restrained NPT equilibration simulations (10 ns in total). In the same fashion, the CheA-trimer and CheA-hexamer subunit models were hydrated and ionized to produce systems of size 1,751,375 atoms (245x245x310 Å) and 4,588,588 atoms (385x405x310 Å) respectively. Each subunit model was then subjected to the same minimization (300,000 steps) and restrained NPT equilibration (10 ns) scheme as the unit cell model. An outline of subsequent equilibration and production simulations is given in Figure 3.10. Production simulations of the unit cell and MDFF-refined CheA-trimer models were conducted with weak (spring constant = 0.1 kcal/mol*nm^2) harmonic restraints placed on the alpha carbons of the first five membrane-proximal receptor residues to maintain TOD splay in the absence of membrane and crowding agents. In the case of the post-MDFF production simulations of the CheA-trimer, additional weak harmonic constraints were placed on the outermost CheW and CheA-P5 domains to enforce the bulk array boundary conditions, as the trimer organization does not permit the use of periodic boundary conditions to represent the necessary symmetry.

All molecular dynamics simulations were performed using the parallel molecular dynamics code, NAMD 2.9 [49] and CHARMM22 force field [96] with CMAP corrections [97]. Equilibrium simulations were conducted in the NPT ensemble with isobaric

and isothermal conditions maintained at 1 atm and 330 K for equilibration, or 350 K for production using the Nose-Hoover Langevin piston, with a period 200 femtoseconds (fs) and relaxation time of 50 fs, and the Langevin thermostat with a temperature coupling of 5 ps^{-1}. The r-RESPA integrator scheme [49] with an integration time step of 2 fs was used. SHAKE constraints were applied to all hydrogen atoms [98]. Short-range, non-bonded interactions were calculated every 2 fs with a cutoff of 12 Å and long-range electrostatics were evaluated every 6 fs using the Particle Mesh Ewald (PME) method [99] with a grid size of 1 Å. Periodic boundary conditions with fixed cross-sectional area (x-y plane) were used. MDFF simulations were performed in the NVT ensemble at 330 K using the settings described above with additional restraints applied to prevent loss of secondary structure, chirality errors, and the formation of cis-peptide bonds.

3.4.2 Simulation analysis

Visualization and extraction of raw trajectory data for analysis were performed using VMD. Principal Components Analysis (PCA) was implemented using routines in the Numpy, Scipy, and MDAnalysis python packages [100]. For the PCA analysis, a single dip-exhibiting CheA dimer was isolated from one of our wild-type unit cell simulations, and each frame (23,331 frames in total) was aligned to the initial CheA dimer model using the P5 domains (residues 543–671). Principal components were computed using the alpha carbons of the P4 domains (residues 352 to 542). The fractional variances accounted for by the top three modes were 41.8%, 31.1%, and 8.1% respectively. Subsequently, the three CheA dimers from each replica of the wild-type unit cell model (27 dimers total), R297A unit cell model (27 dimers total), and CheA-trimer model (30 dimers total) simulations were extracted, aligned to the P5 domains, and projected on to the top principal component of the wild-type dip-exhibiting CheA dimer. These projections were grouped according to model type to create Figure 3.12. Illustrations of the PCA results were pro-

duced using the python-plotting package, Matplotlib [101].

Clustering analysis routines were implemented using the python packages noted above as well as the UPGMC hierarchical, agglomerative clustering function from the fastcluster package [95]. For the clustering analysis, we first extracted the three core signaling units from each of the nine wild-type unit cell replica simulations, using 1500 frames/core signaling unit for a total of 40,500 frames. The RMSD distance matrix was then computed using the QCP structure alignment function from the MDAnalysis package. Our analysis identified four major clusters of structures within the above distance matrix with relative populations of 80%, 10%, 10% and 2%, representing the "undipped", "dipped", and two intermediate CheA dimer states respectively.

3.4.3 Experimental methods

A discussion concerning the experimental methodologies used in this study can be found in [2], including (1) protein expression and purification, (2) monolayer reconstitution, (3) cryo-electron tomography, (4) 3D reconstruction, sub-tomogram classification and averaging, (5) mutagenesis, (6) cross-linking and western blot analysis, and (7) soft agar assays.

3.5 Supplemental Information

Table 3.1: Inter-protein interactions at key interfaces of *T. maritima* unit cell model.

Interface	Domain	Residues	References
CheA-P5/ Receptor	CheA-P5	L547, I560, I563, I566, L629	23, 24, 25, 27, 28
	Receptor	L362, L365, N366, A368, I369, A372	
	CheA-P5/ Receptor	D564/R373(1)	
CheW/ Receptor	CheW	L14, V27, I30, V33, V98, V101	23, 24, 25, 26, Present study
	Receptor	L362, L365, N366, A368, I369, A372	
	CheW/ Receptor	K9/R376(1,3), E10/R379(1,3)*, E12/R379(2)*, D28/R373(2)*, E31/R373(2)*	
CheA-P5/ CheW interface I **	CheA-P5 (subdomain 1)	L554, L599, I601, L640, V643, F644, V647, F650, A653, I655, I661, L663	20, 23, 24, 29, 30
	(subdomain 1)	V647, F650, A653, I655, I661, L663	
	CheW (subdomain 2)	T40, P41, V42, P43, V49, V52, I59, P61, V63, V89, V91, I145	
	CheA-P5/ CheW	K642/D88, E646/R46, E649/K44,	
CheA-P5/ CheW interface II **	CheA-P5 (subdomain 2)	Q575, V577, Q578, V582, V584, V589, P591	23, 24
	CheW (subdomain 1)	T112, N113, V114, S115, F118, L125, L132, I134	
	CheA-P5/ CheW	R580/D116, E588/K127	
CheW/ CheW	CheW (subdomain 1)	Q21*, L69*, I71*, T112*, N113*, V114*, S115*, F118*, L125*, L132*, I134*	Present study
	CheW (subdomain 2)	T40*, P41*, V42*, P43*, V49*, V52*, I59*, P61*, V63*, V89*, V91*, I145*	
	CheW/ CheW	D116/R46*	
Receptor/ Receptor	dimer/ dimer	E370-R373, E351-R403*, E387-R389*	18, Present study

Summary of residues participating in a given interface but not associated with particular partners are listed separately for each domain. Residues that interact significantly (>50% of frames) are listed as a pair in a separate row. Interactions unique to this study are signified with *. Where ambiguous, residue pairs involving a receptor bound to (1) CheA-P5, (2) CheW from a CheA-P5/CheW ring or (3) CheW from a CheW-only ring are numbered as denoted here. Interfaces taken directly from experimental structures are signified with **. Recent references pertaining to each protein-protein interface are given.

Table 3.2: Summary of key inter-protein interface changes between "dipped" and "undipped" CheA conformations.

Interface	Domain	Residues	References
CheA-P3/Receptor (Undipped)	CheA-P3/ Receptor	D333/K390(2)*, D345/R379(2)*	25, Present study.
CheA-P3/CheA-P4 (Undipped)	CheA-P3/ CheA-P4	K352/E390, R354/D392, D304/R393*	19, Present study.
CheA-P4/Receptor (Dipped)	CheA-P4/ Receptor	E390/R379(2)*	Present study.
CheA-P3/CheA-P4 (Dipped)	CheA-P3/ CheA-P4	K352/E390, R354/D392, R297/E397*	19, Present study.

Interactions unique to this study are signified with *. Where ambiguous, residue pairs involving a receptor bound to (1) CheA-P5, (2) CheW from a CheA-P5/CheW ring or (3) CheW from a CheW-only ring are numbered as denoted here. Recent references pertaining to each protein-protein interface are given.

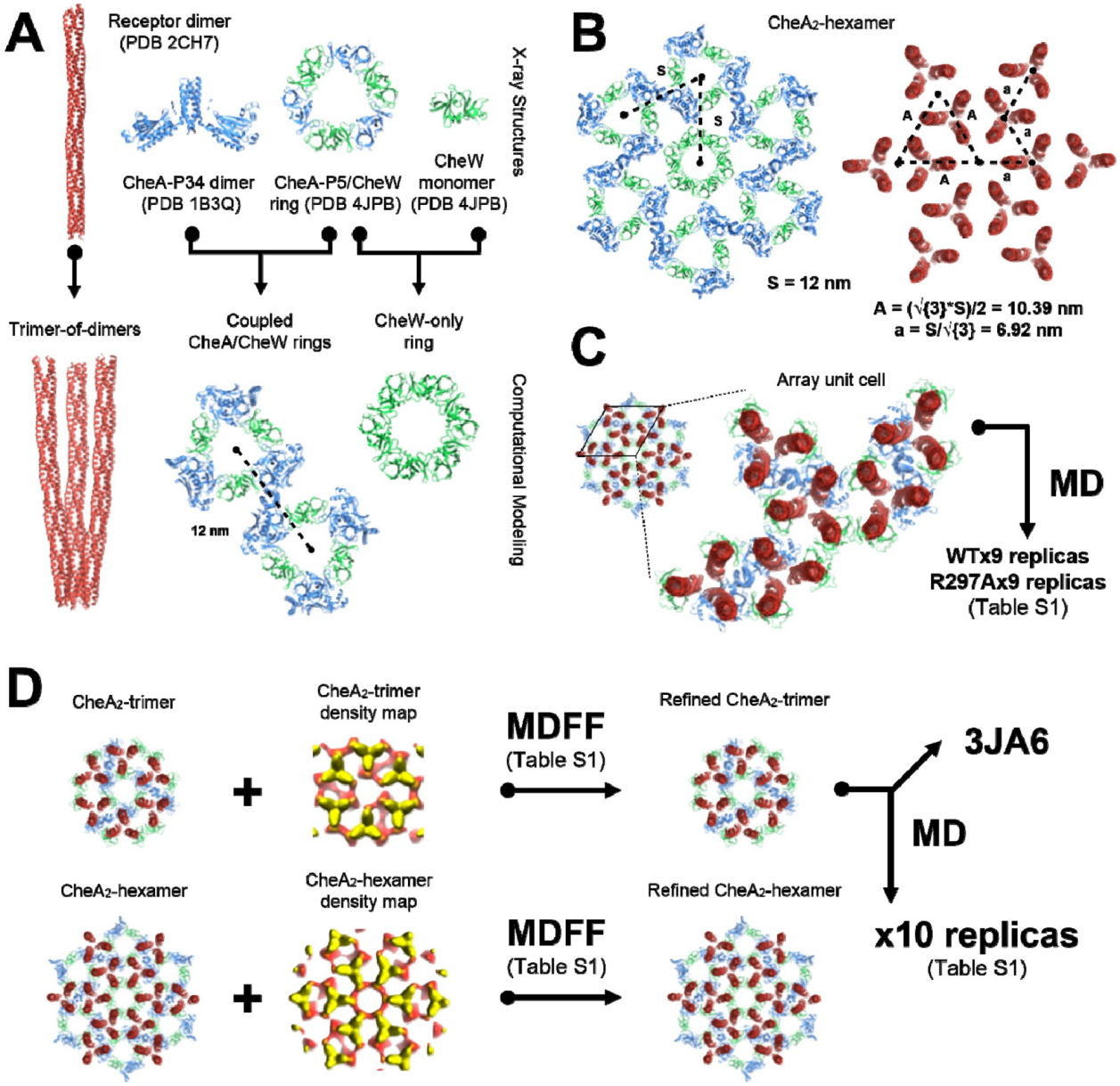

Figure 3.15: (A) High resolution X-ray structures from *T. maritima* were taken as inputs for the generation of models corresponding to the array's core components, namely the receptor trimer-of-dimers, coupled CheA/CheW rings, and CheW-only ring. (B) The resulting core component models were arranged heuristically, assuming a 12 nm lattice constant, to produce models of the CheA-hexamer and CheA-trimer array substructures. For simplicity, only the CheA-hexamer organization is shown. (C) A portion of the heuristically-constructed CheA-hexamer model corresponding to the array unit cell was extracted for further study with all-atom MD simulations. (D) MDFF simulations were conducted to refine the CheA-hexamer and CheA-trimer models utilizing their respective density maps. A core signaling unit was taken from the MDFF-refined CheA-trimer model and deposited in the PDB data bank under accession code 3JA6. The full MDFF-refined CheA-trimer model was subjected to further investigation using all-atom MD simulations.

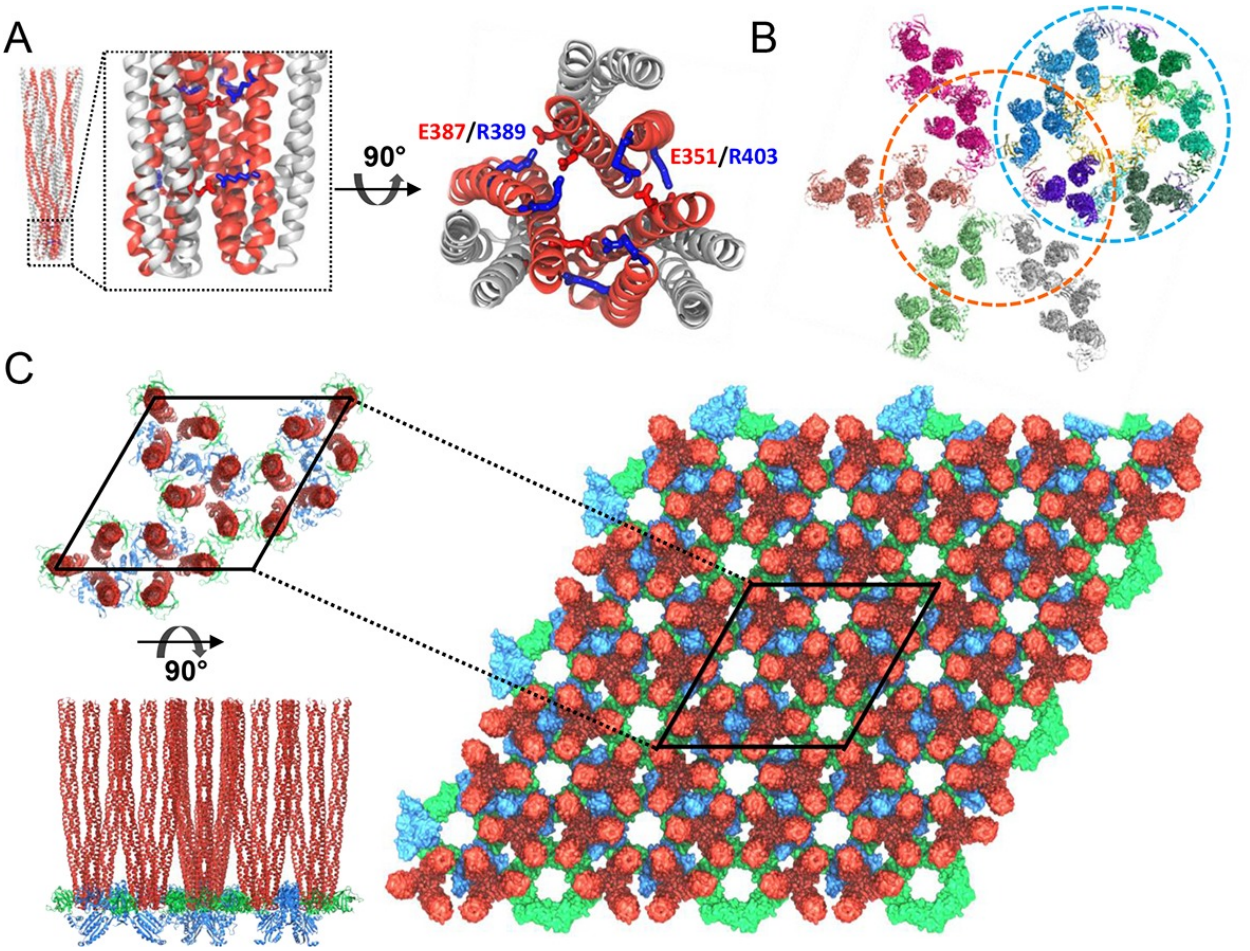

Figure 3.16: Computational modeling of the extended chemosensory array structure. (A) Molecular dynamics (MD) simulations show that *T. maritima* receptors form a stable trimer-of-dimers (TOD). Side view (left) and top view (right) of the highly conserved protein interaction tip, highlighting the inter-receptor salt bridge network formed by E351/R403 and E387/R389. Symmetry-related monomers within individual receptor dimers are distinguished by red and grey coloring. (B) MDFF-refined, all-atom model of the array subunits combining CheA2-trimer (cyan circle) and CheA2-hexamer (orange circle) maps. (C) All-atom model of the *T. maritima* lattice containing 3 x 3 unit cells. A portion of the lattice corresponding to a single unit cell is outlined in black. Top (top left) and side (bottom left) views of the array unit cell model arranged as three coupled core-signaling units. Receptor TODs are shown in red, CheA dimers in blue, and CheW monomers in green.

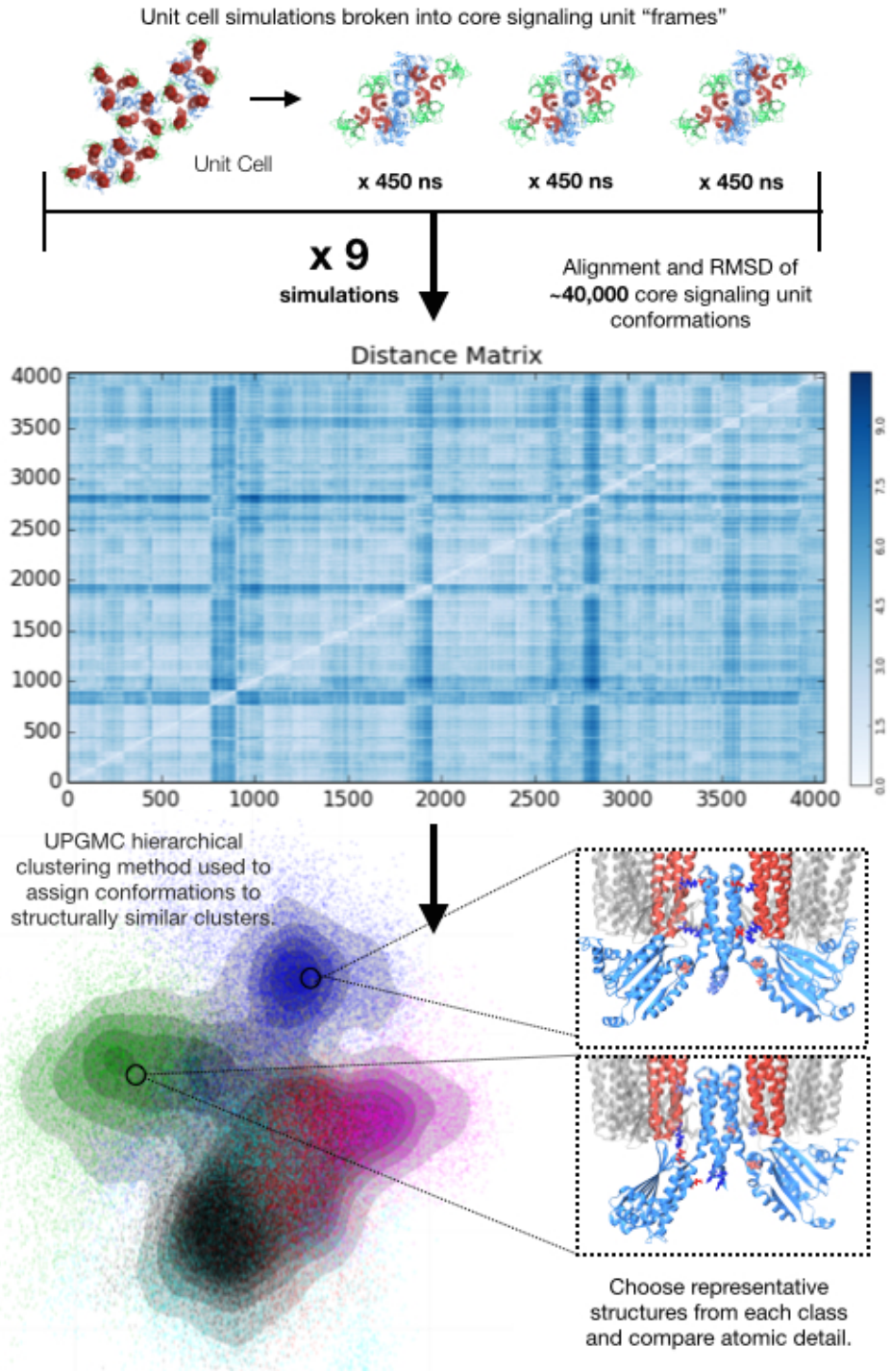

Figure 3.17: UPGMC hierarchical clustering analysis of wild-type unit cell simulations. We first divided each of the nine wild-type unit cell replica simulations (450 ns = 1500 frames) into three core signaling units, totaling 9x1500x3 = 40,500 frames. We then computed the RMSD between each frame to construct the pairwise distance matrix. The UPGMC hierarchical clustering method was then use to assign each frame to clusters in an agglomerative fashion. Our analysis identified classes of structures differing significantly in structure, which could be individual analyzed for unique features. CSUs from our R297A unit cell and CheA-trimer simulations were clustered in an identical fashion.

Chapter 4

The *E. coli* Serine Receptor: Modeling and Simulations

4.1 Introduction

The chemotaxis machinery of the model bacterium *E. coli* is the best understood sensory signal transduction system in biology. It has served as a powerful experimental model for investigating the molecular mechanisms that proteins use to detect, process, and transmit stimulus information [24, 102–104]. Central to the signal transduction underlying chemotactic behavior in *E. coli* are dedicated chemoreceptors (also known as methyl-accepting chemotaxis proteins or MCPs), which couple ambient concentration gradients to the activity of an associated histidine autokinase (CheA) and ultimately cellular swimming pattern [30, 31, 40]. Studies of two *E. coli* MCPs, in particular, Tsr and Tar (attracted to serine and aspartate respectively) form the foundation of the current understanding of MCP structure and signal transduction mechanisms. In this section, I will provide an overview of those topics.

As depicted in Figure 4.1, MCPs are homodimeric, transmembrane proteins that form an $\sim$35-nanometer-long series of coiled-coils and helices. MCPs can be divided into three functionally distinct signaling regions: (1) stimulus sensing, (2) input/output control, and (3) kinase control. The stimulus-sensing region is comprised of two modules, a periplasmic ligand-binding domain and a transmembrane (TM) four-helix bundle. Through the binding of chemical stimuli, conformational changes are promoted within the stimulus-sensing region that allow the transduction of sensory signals across the cell's inner mem-

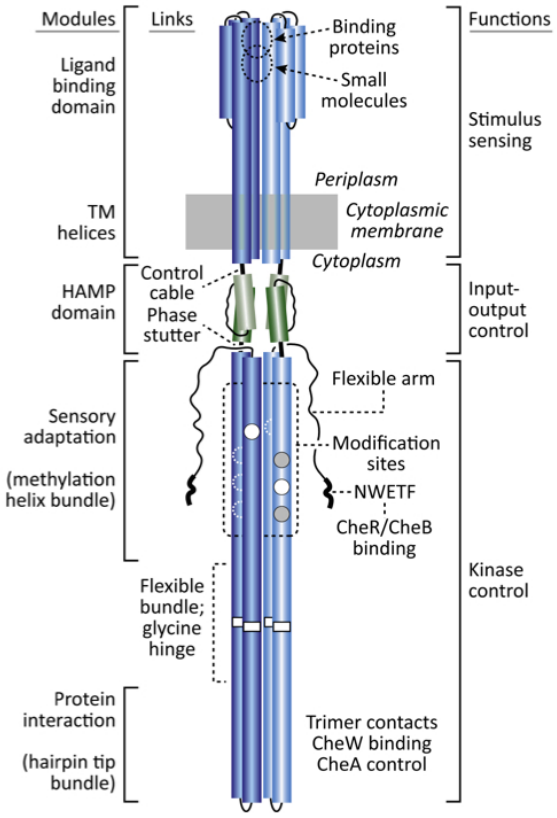

Figure 4.1: Overview of the structural and functional organization of a methyl-accepting chemotaxis protein (MCP). The two monomers of the MCP homodimer are shown in different shades of blue, the HAMP domain is highlighted in shades of green. Helical regions are shown as cylinders. On the left are listed the individual MCP modules, which can be more broadly grouped into functional regions listed on the right. The sites of other key functional features are listed, including those of ligand-binding, adaptational modification, trimer-formation, and CheR/CheB binding. Figure reproduced with permission from [31].

brane via a piston mechanism [105, 106]. The stimulus-sensing and input/output-control regions are connected by a five-residue linker, the so-called control cable [107–109], which subtly mediates the transduction of signals to the HAMP domain, a ubiquitous signaling module found in Histidine kinases, Adenylyl cyclases, MCPs, and Phosphatases [110–114]. The HAMP domain is, in turn, connected to the kinase-control region via a four-residue "phase stutter" that gives rise a structural displacement affecting the packing of the surrounding coiled-coils [115]. The kinase-control region, as its name suggests, regulates the autophosphorylation activity of the associated CheA histidine kinase. It contains two modules: the sensory-adaptation module and protein-interaction module, which are

separated by a flexible glycine hinge [116]. The latter module contains the contacts which directly bind CheA and CheW [74,76,77] as well as those which participate in the formation of the trimer-of-dimers (TOD) [69,117], enabling the assemblage of the highly cooperative extended array architecture discussed in Chapters 1 and 3. The sensory-adaptation module is used to tune the detection sensitivity of MCP molecules through the reversible methylation of specific glutamyl residues (i.e., adaptation sites) [28, 29]. Finally, the C-terminus of the MCP contains a flexible arm that is concluded with an NWETF motif, which binds CheR and CheB to enable efficient adaptational modification of MCPs [118].

Extensive genetic, biochemical, biophysical, and computational studies have produced an intriguing view of signaling events within the cytoplasmic portion of MCPs [31, 32, 112, 114, 119–121]. Although signaling within core signaling units (CSUs) approximates two-state behavior, involving ligand and methylation-driven shifts in an equilibrium between complexes with kinase-off and kinase-on outputs, the underlying structural elements of individual MCP molecules appear to operate over a series of essentially isoenergetic conformational states. In the above picture of MCP signal transmission, shown in Figure 4.2, adjacent signaling elements are coupled in opposition, such that a stabilizing shift in one element is offset by a destabilizing shift in its neighbors. These (de)stabilizing shifts are thought to arise from changes in the strength of inter-helical packing within each signaling element, leading to either a more static (stable packing) or more dynamic (unstable packing) state. In particular, it has been proposed that the kinase-on and kinase-off states are respectively characterized by static-dynamic-static and dynamic-static-dynamic couplings between the HAMP, sensory-adaptation, and protein-interaction modules. This molecular communication strategy probably poises the signaling tip of receptor molecules in an intermediate dynamic state that can be readily shifted toward more static or more dynamic behavior by low-energy stimulus inputs such as binding of a small ligand molecule.

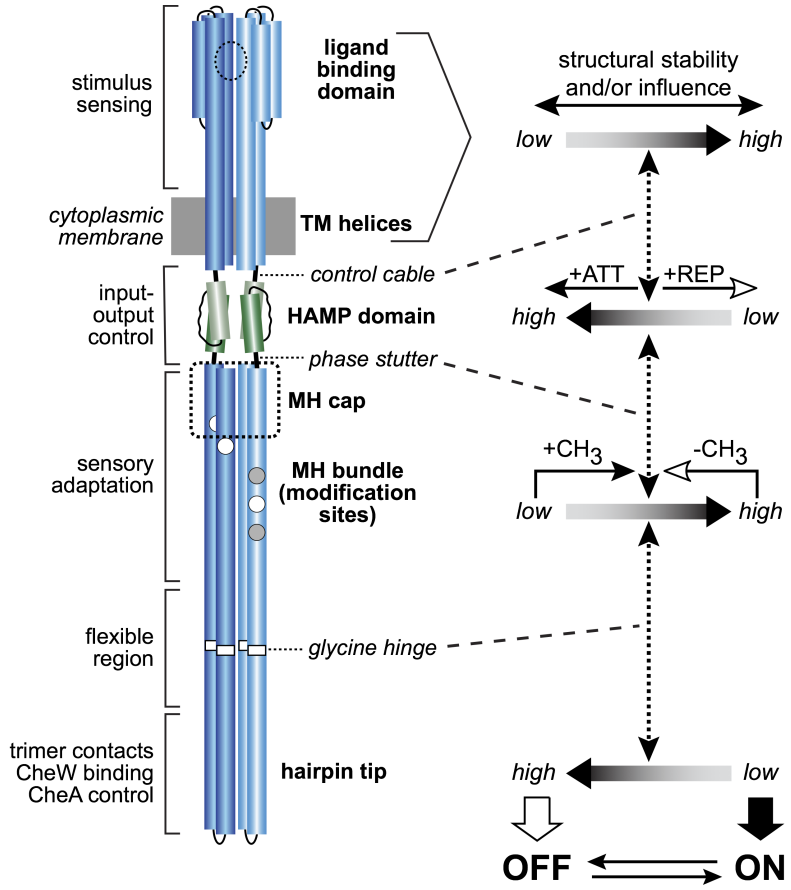

Figure 4.2: Static-dynamic model of MCP signaling. Adjacent signaling elements are coupled in opposition, such that a stabilizing shift in one element is offset by a destabilizing shift in its neighbors. The particular pattern of stability, as dictated by ligand occupancy and methylation state, gives rise to the overall state of the MCP, favoring either the kinase-on/off state. Figure provided by Sandy Parkinson, U. Utah.

Though the above picture represents the culmination of a substantial amount of data, the detailed molecular events giving rise to particular alternating static-dynamic patterns as well as the ways in which these patterns are influenced by chemical stimuli and adaptational modifications are still quite mysterious. Although a number of crystallographic structures exist for portions of MCPs from varying species [69, 71, 122–125], a high resolution picture of the complete MCP structure has remained elusive, hindering the investigation of the function of MCPs and their higher-order organizations. In this chapter, I describe my work to construct an atomic model of the intact, membrane-bound *E. coli* serine receptor.

4.2 Results

4.2.1 Atomic model and simulations of Tsr ligand-binding, transmembrane, and HAMP domains

The *E. coli* serine receptor (Tsr) spans the inner membrane using two helices, denoted TM1 and TM2, which form a transmembrane (TM) four-helix bundle in the MCP homodimer (Fig. 4.3). Considerable experimental [105, 106, 126, 127] and computational [128,129] evidence suggests that ligand binding induces a ∼2 Å piston-type motion in TM2 (Fig. 4.3) [105]. Previously, it has been assumed that this piston motion is propagated directly through the membrane to the HAMP domain and further downstream [106].

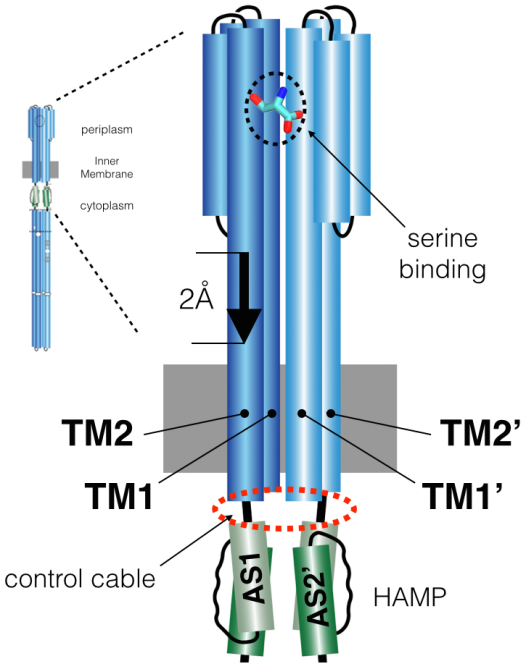

Figure 4.3: Transmembrane signaling in *E. coli* serine receptor (Tsr). The Tsr transmembrane (TM) domain is a four-helix bundle. Each monomer contributing two helices to the bundle, denoted TM1 and TM2. Serine binding (black circle) is thought to evoke a ∼2 Å piston-type motion in TM2 that is transmitted across the cell membrane. A five-residue control cable (red circle) mediates TM signals between TM2 and the AS1 helix of the HAMP domain via an unknown mechanism.

Recently, studies conducted by our collaborator Sandy Parkinson (University of Utah) have given rise to a more complex transmembrane signaling picture, involving a key five-residue "control cable" region [107,109]. In particular, exhaustive mutagenesis of the Tsr control cable has demonstrated that alterations to this area can dramatically affect signaling behavior (Fig. 4.16, Supplemental Information) and suggests that control cable helicity may mediate transitions between the kinase-ON and kinase-OFF signaling

states [107]. In addition, they show that a single residue (I214) is largely responsible for arbitrating changes in signaling state [109]. To explain their extensive signaling mutant data, Parkinson *et al.* developed a model in which the kinase-ON state is predicted to have a highly helical control cable that exacerbates the helical register mismatch between TM2 and HAMP AS1, while the kinase-OFF state is predicted to have a reduced helical content that relieves this mismatch. However, a high-resolution structural description of the TM four-helix bundle and control cable region is not available. Hence, while it is clear that the experimentally observed behavior can not be explained using a simple piston model alone, the structural and dynamical information needed for a full, molecular interpretation of the experimental results is lacking.

To address this deficiency, we have utilized MD simulations, in tandem with biochemical and crystallographic data, to construct an atomic model of the Tsr TM four-helix bundle together with the ligand-binding and HAMP signaling domains. Specifically, we have made use of existing disulfide cross-linking efficiencies [126, 130] from the *E. coli* aspartate receptor (Tar) to set bounds on the distances between specific residues in the highly conserved Tsr TM four-helix bundle (Fig. 4.15, Supplemental Information). Shown in Figure 4.13 and Figure 4.14 (Supplemental Information), these data highlight specific residues at the inter-monomer (TM1/TM1', TM1/TM2', TM2/TM2') and intra-monomer (TM1/TM2) interfaces, that readily form (or do not form) disulfide cross-links when mutated to cysteine. As cross-linking requires that the mutated residues come into close proximity to one another (Cβ-Cβ distance $<\sim$12 Å), and the efficiency of crosslinking is a function of distances between the residues, these data can be used as a guide for informing the rough placement (i.e., helical tilt and register) of the four TM helices. Combined with high-resolution information of the upstream and downstream structural regions (i.e., ligand-binding and HAMP domains respectively), we were able to obtain an atomic model that agreed well with existing experimental results as described below.

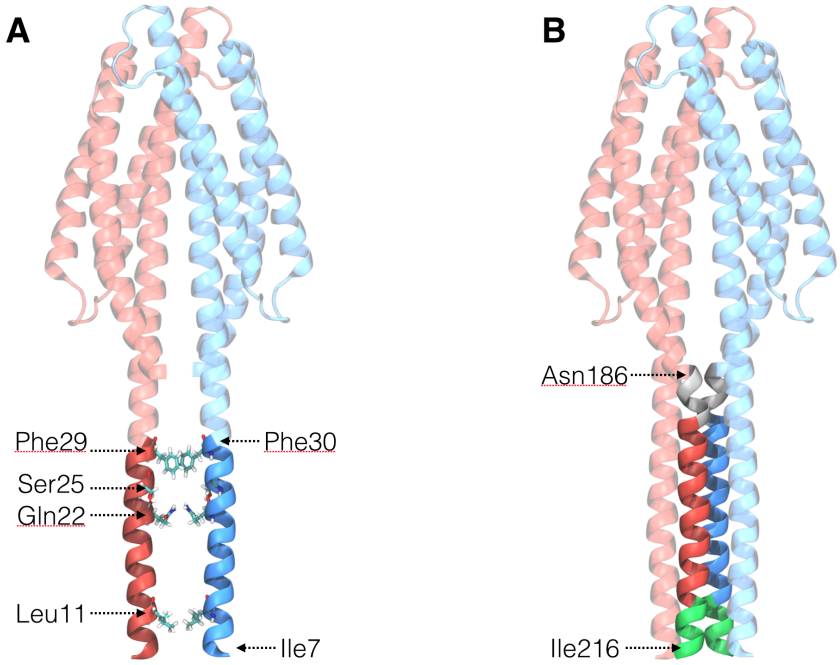

Figure 4.4: *E. coli* Tsr ligand-binding domain with modeled TM and control cable helices. (A) APO crystal structure of the Tsr periplasmic domain (PDB 2D4U, residues 31-185), shown in reduced opactiy, with modeled TM1 and TM1' helices (residues 7-30) shown with full opacity. Residues at the TM1/TM1' interface that efficiently cross-link are labeled and shown in licorice. (B) Addition of TM2 (opaque red and blue) and control cable (green) helices. A short segment of connecting helix (gray) was modeled between PDB 2D4U and TM2 helices.

Our modeling approach proceeded in several steps. First, as disulfide cross-linking studies have demonstrated a strong TM1/TM1' interface that is unaffected by signaling [106, 126], we began by modeling the TM1 helices (residues 7-30) assuming ideal alpha helices [106]. Using a crystal structure of a portion of the *E. coli* Tsr periplasmic domain (PDB 2D4U, residues 31-185) [122], we extended the structure to include TM1 and TM1' (Fig. 4.4.A). To obtain an energetically favorable side chain packing between the TM1 helices, we conducted a 500 ns simulation of the Tsr periplasmic+TM1 model. In addition, to drive our MD simulation toward a packing consistent with the experimental data, we used the existing disulfide crosslinking data on the TM1/TM1' interface (Fig. 4.13) to construct a series of collective variable potentials within NAMD [49]. In particular, we chose symmetric residue pairs that were shown to cross-link with either high

or detectable efficiency (i.e., residues 7, 11, 14, 15, 18, 19, 21, 22, 25, 29, 33) and constructed wall potentials such that if the Cβ-Cβ distance between residues was less than 12 Å no bias was exerted, otherwise a harmonic spring acted to bring the pair within the specified range. After obtaining a stable TM1/TM1' interface, we added the TM2 helices (residues 191-210) in a similar fashion. Moreover, as mutagenesis evidence suggests that the control cable (residues 211-215) is also likely alpha helical in nature, forming a continuous helix between TM2 and AS1 in at least some native signaling states [107, 109], we added this region as well (Fig. 4.4.B). To obtain an appropriate side chain packing, we ran an additional 500 ns simulation of the Tsr periplasmic+TM1+TM2 model, utilizing the existing disulfide crosslinking data on the TM1/TM2 and TM1'/TM2 interfaces (Fig. 4.14) [130] to construct "wall" potentials as described above.

To provide cytoplasmic structural constraints on the TM1 and TM2 helices we constructed models of their respective downstream structural elements. In the case of TM1, the short N-terminus (residues 1-6) was predicted by the JPred4 server [131] to be a random coil, in line with its lack of participation in TM signaling [132]. Hence, we used the program Modeller [133] to add an appropriate, non-clashing coil to the equilibrated Tsr periplasmic+TM1+TM2 model. In the case of TM2, each helix is connected (via the control cable) to the HAMP domain (Fig. 4.3). However, no structure exists for a HAMP domain from any bacterial MCP. Nevertheless, a crystal structure of a highly conserved (Fig. 4.5.A) archaeal HAMP domain from the Af1503 receptor of *Archaeoglobus fulgidus* (PDB 2L7H) is available [113]. Hence, we used Modeller [133] to generate a homology model of the Tsr HAMP four-helix bundle (Fig. 4.5.B, residues 216-268) using PDB 2L7H as a structural template. We then used the CHARMM-GUI program [134], to place the protein in a 3:1 POPE:POPG lipid bilayer, mimicking the inner membrane of *E. coli* [135]. The resulting model, hereafter referred to as Tsr_{TM} (Fig. 4.5.C), was then subjected to a 1 microsecond, unbiased equilibration simulation.

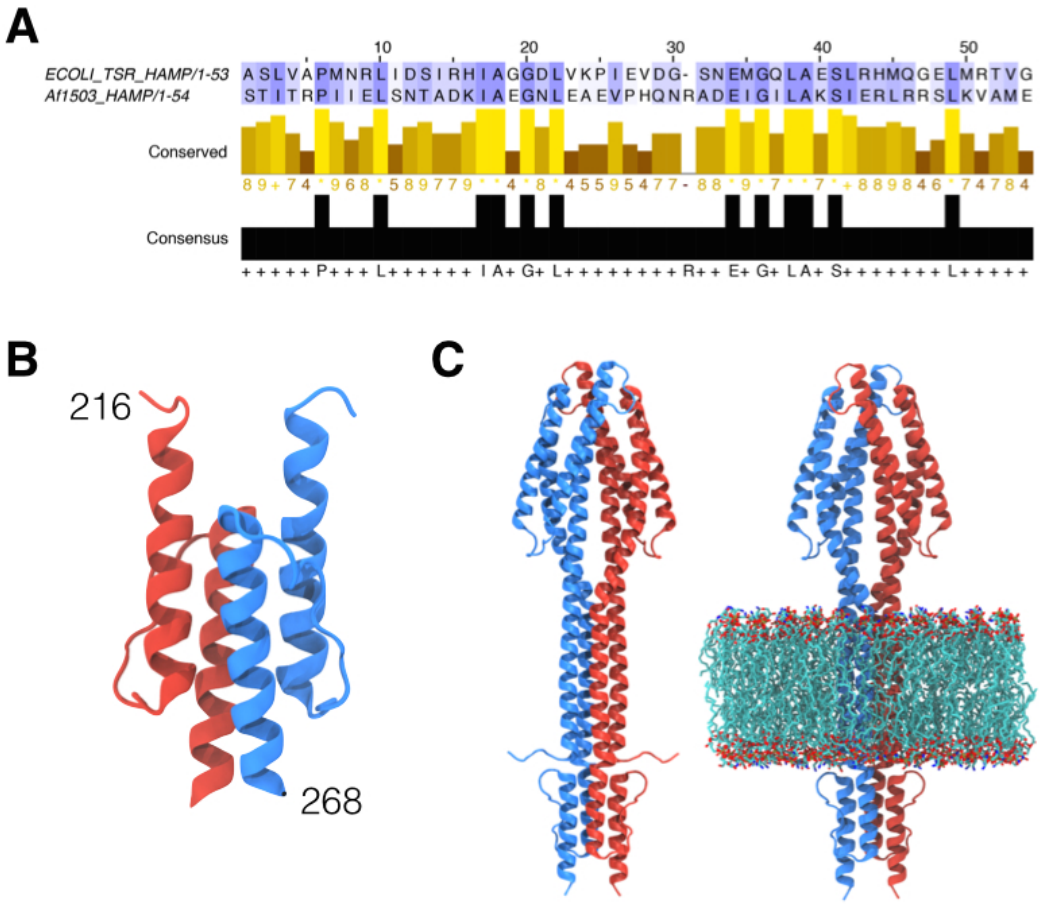

Figure 4.5: (A) MAFFT pairwise sequence alignment between *E. coli* Tsr and archaeal Af1503 HAMP domains. Alignment colored using BLOSUM62 score [136]. The conservation of each residue and consensus sequence (showing identical residues) are shown directly below the alignment. Panel created using Jalview [137]. (B) Homology model of the Tsr HAMP domain (residues 216-268) (C) Intact, Tsr_{TM} model, including periplasmic, TM four-helix bundle, and HAMP domains with (right) and without (left) POPE:POPG lipid bilayer. Receptor monomers are shown in red and blue.

The equilibrated Tsr_{TM} model agreed well with previous experiments. In particular, the unconstrained TM four-helix bundle was seen to form a stable packing, which nicely corresponded with our interpretation of the disulfide cross-linking efficiencies (Fig. 4.13, Fig. 4.14, Supplemental Information). Figure 4.6 shows the mean Cβ-Cβ distance (as computed over the course of the one microsecond equilibration simulation of Tsr_{TM}) between symmetric residues along the two TM1 (panel A) and two TM2 (panel B) helices. Residue pairs (converted to Tsr numbering) that were shown to cross-link with high (green) and

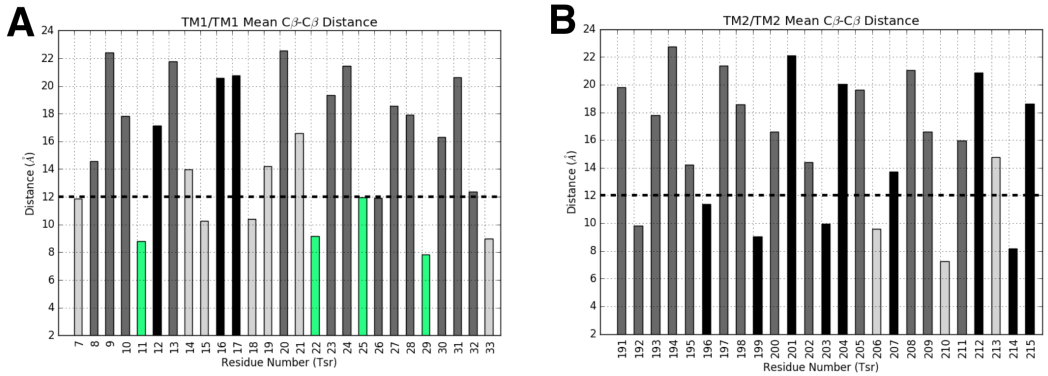

Figure 4.6: Mean Cβ-Cβ distances between symmetric residues in the (A) TM1/TM1′ and (B) TM2/TM2′ interfaces, computed over the course of a one microsecond equilibration simulation of Tsr_{TM}. Residues are numbered according to *E. coli* Tsr. The coloring of residue pairs is based on Tar cross-linking data (Fig. 4.13 [130]. Green and light gray denote pairs that were shown to cross-link with high or intermediate efficiency respectively. Dark gray denotes pairs that did not detectably cross-link, and black denotes pairs for which there was no information provided. A dashed line, drawn at 12 Å, denotes rough distance necessary for cross-link formation. (Note, for Tsr residue numbers less than 84 (i.e., TM1), Tsr and Tar have the same numbering, while for Tsr residue numbers greater than 116 (i.e., TM2), Tsr = Tar + 2.)

intermediate (light gray) efficiencies are largely separated by less than 12 Å, while residue pairs that did not show detectable cross-linking (dark gray) are largely separated by more than 14 Å. It should be noted, though, that due to the nature of cross-linking experiments, fluctuations in proteins may give rise to cross-links that trap the protein in sparsely occupied or non-functional conformations [138]. Hence, strict adherence to the cross-linking data is not expected. In addition, Figure 4.7 shows the mean Cβ-Cβ distance between pairs of residues at the intra-molecular (panel A) and intermolecular (panel B) TM1/TM2 interfaces. Curiously, at these interfaces, a number of reported cross-links were not unambiguously identified (Fig. 4.14). Each such pair is marked with a particular solid black shape. In all cases but one (21/203 vs. 21/205), our model clearly resolves a spatially favored cross-link (blue vs. gray bars for each black marker shape). In the exceptional case, examination of our equilibrated model shows that, although, residues 21 and 203 are in fact near one another, cross-linking would require a rotation of the TM2 helix that would potentially distort the TM four-helix bundle. Indeed, the analogous cross-link in the homologous *Salmonella typhimurium* Tar was shown to destroy signaling [139].

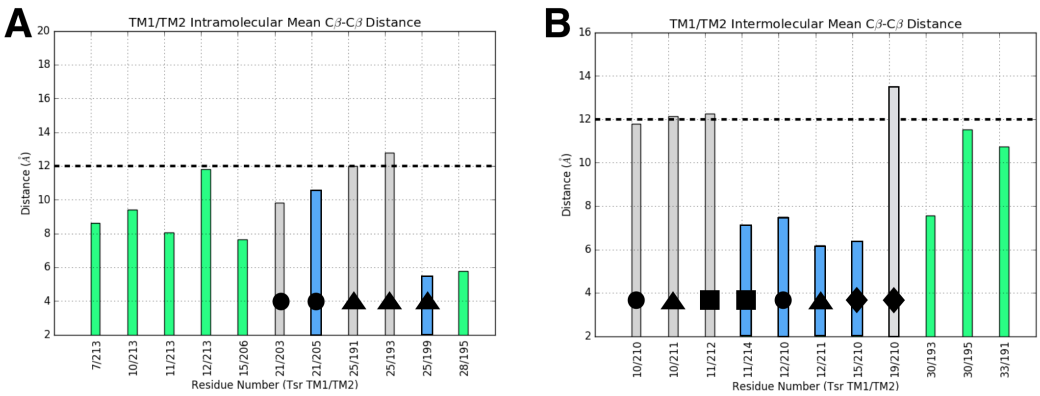

Figure 4.7: Mean Cβ-Cβ distances between key residues in the (A) TM1/TM2 and (B) TM1/TM2′ interfaces, computed over the course of a one microsecond equilibration simulation of Tsr_{TM}. Residues are numbered according to *E. coli* Tsr. The coloring of residue pairs is based on cross-linking data in Tar (Fig. 4.14) [130]. Green denotes unambiguous pairs of residues that cross-linked with high efficiency. Black shapes (circle, triangle, square, diamond) mark pairs where residues could not be unambiguously identified. Within each such pair, a blue bar denotes the cross-link favored by the equilibrated model; gray bars denoted unfavored pairs. A dashed line, drawn at 12 Å, denotes rough distance necessary for cross-link formation. (Note, for Tsr residue numbers less than 84 (i.e., TM1), Tsr and Tar have the same numbering, while for Tsr residue numbers greater than 116 (i.e., TM2), Tsr = Tar + 2.)

The equilibrated Tsr_{TM} system also highlights key protein/lipid interactions (Fig. 4.8.A). In particular, our model correctly positions the TM2 helices relative to the lipid bilayer, situating Y189 and A216 just above and below the periplasmic and cytoplasmic membrane-water interfaces respectively, as previously demonstrated in Tar [140]. In addition, several lysine (K) residues lie at either membrane-water interface that hydrogen bond strongly with the polar lipid head regions, including K3 and K6 in the unstructured N-terminus tail as well as K215 and K34 directly adjacent to Y189 and A216 respectively. Hence, while the N-terminus may not participate directly in TM signaling [132], it is likely quite important for appropriately anchoring the receptor within the lipid bilayer. Finally, two tryptophan (W) residues, W194 and W211, are seen to lie buried roughly 1.5 helical turns within the membrane. This distance is consistent with their ability to modulate TM signaling via an induced piston-like displacement on the order that arising from ligand binding [141]. In general, tyrosine, tryptophan, and lysine are all known to be important for precisely positioning transmembrane helices [142–145].

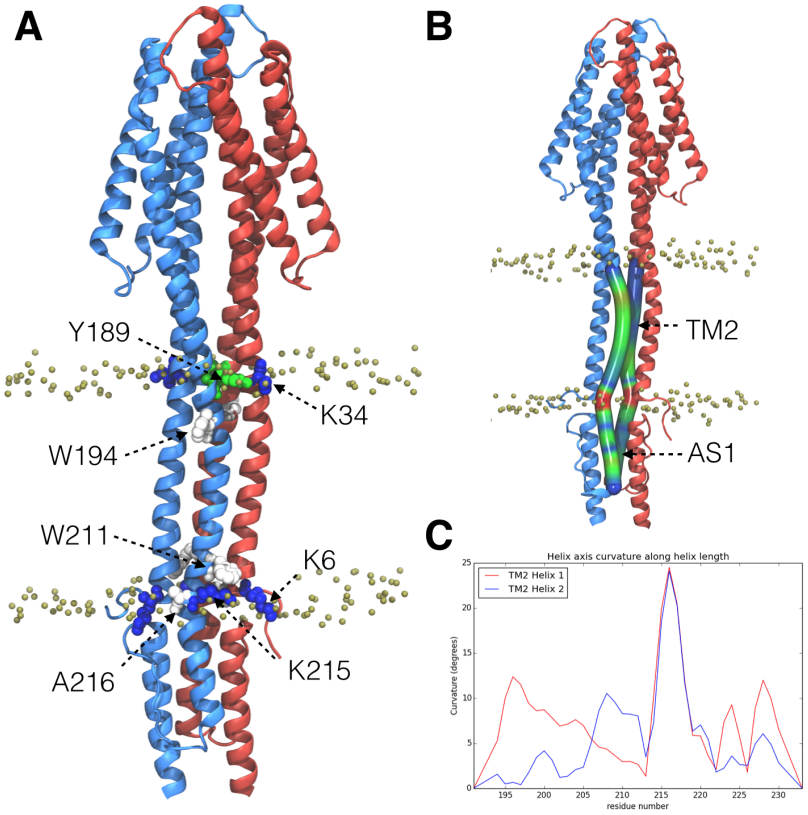

Figure 4.8: Protein/lipid interactions in equilibrated Tsr_{TM} model. (A) Key interactions at periplasmic and cytoplasmic membrane-water interfaces. Residues are shown in licorice and colored according to their electrostatic properties: positively charged (blue), polar (green), and hydrophobic (white). Lipid headgroups are shown as gold spheres. (B) Control cable kink between TM2 and HAMP AS1 at the cytoplasmic membrane-water interface. Helices are colored by helical axis curvature using a BGR scale. (C) Plot of helical axis curvature along TM2, control cable, and HAMP AS1 helices (residues 191-235), showing a ($\sim 25^o$) kink (as measured by VMD plugin bendix [146]) at A216 in the control cables of both MCP monomers (red and blue traces).

Our atomic model also shed light on the detailed structure of the elusive control cable linker region (residues 213-217). Overall, the control cable is alpha helical (Fig. 4.11.A). Nevertheless, as depicted in Figure 4.8.B and Figure 4.9.B, the TM2, control cable, HAMP AS1 helices do not share a continuous axis, but are rather separated by a $\sim 25^o$ kink in the control cable at A216 near the membrane-water interface (Fig. 4.8.C). Two other control cable residues, namely G213 and S217, form a high-occupancy backbone hydrogen bond that partially stabilizes this kink (Fig. 4.9.B&C). K215, as previously described, is located at the cytoplasmic membrane-water interface and is oriented such that its sidechain

projects outward into the lipid bilayer (Fig. 4.9.A), allowing interactions with the surrounding lipid headgroups. Interestingly, the key I214 "trigger" residue is buried just within the membrane with its sidechain directed inward towards the TM1/TM1′ interface (Fig. 4.9.A) and situated within a tight, hydrophobic pocket created by V8 and L12 of the neighboring TM1′ helix.

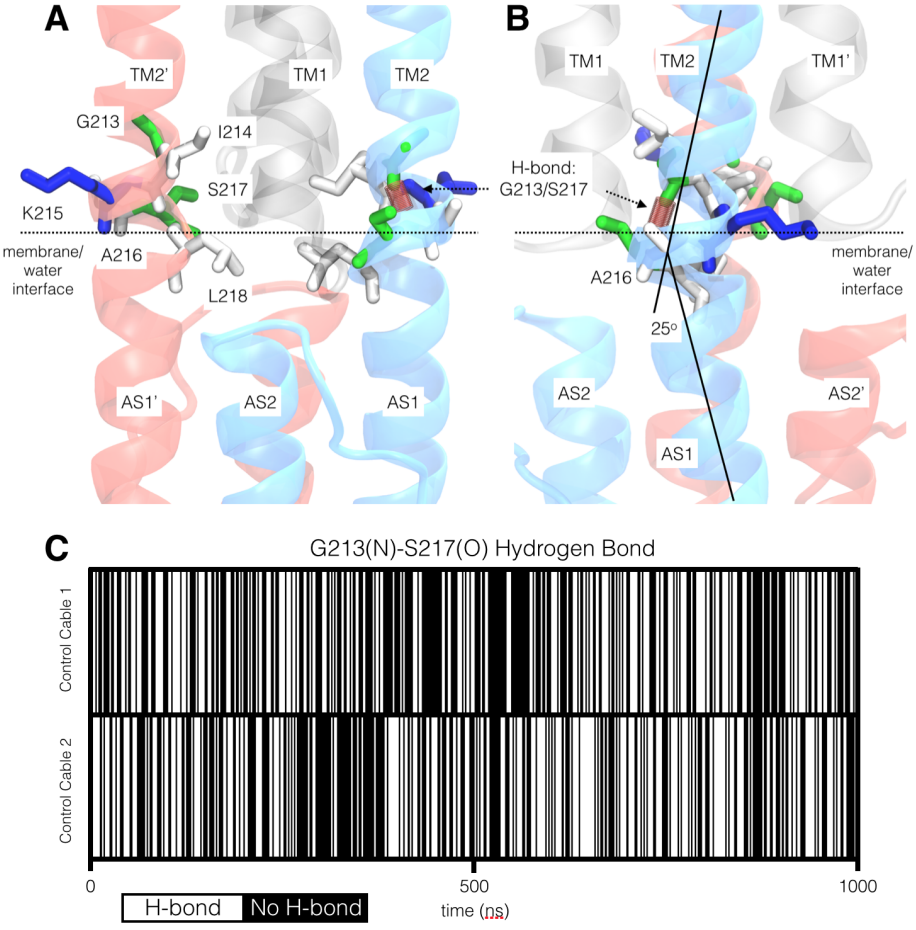

Figure 4.9: Atomic structure of control cable region. (A) Resdiues 213-218 are shown in licorice and colored according to their electrostatic properties: positively charged (blue), polar (green), and hydrophobic (white). I214, a key "trigger" residue, is buried just within the membrane surface with it's sidechain oriented inward towards the symmetry axis of the TM four-helix bundle. A highly occupied backbone hydrogen bond between G213 and S217 is depicted as a red spring. The helices of each MCP monomer (red and blue) are labeled accordingly. For clarity, the membrane-water interface is shown as a dashed line and TM1 helices with reduced opacity. (B) Same as panel A, rotated 90^{o} clockwise as viewed from top, illustrating A216 kink. (C) Occupancy of G213/S217 backbone hydrogen bond for over course of one microsecond equilibration simulation. Frames in which the H-bond is present in control cable 1 (top, blue in panel A) and control cable 2 (bottom, red in panel A) are shown as white bands.

4.2.2 I214 mutations modulate control cable helicity

To investigate the dynamics of the control cable region and the dramatic effect of alterations to I214 witnessed by Parkinson *et al.* (Fig. 4.16, Supplemental Information), we implemented two mutations *in silico*, namely I214A and I214D, that were shown to have opposite signaling phenotypes. In particular, the I214A receptor was seen to be shifted towards the kinase-ON state, while the I214D receptor was shifted toward the kinase-OFF state. We then conducted a one-microsecond, all-atom MD simulation of each mutant. Overall, the mutant simulations closely resembled the wild-type model in the ligand-binding and TM four-helix bundle regions. Moreover, in the I214A simulation, the control cable was observed to nicely maintain its alpha helical nature (Fig. 4.11.B). However, in the I214D simulation, we observed a significant change in the helical content of one of the control cables (Fig. 4.11.C). Specifically, the mutation of I214 to a positively charged aspartic acid residue led to the ejection of both side chains from the hydrophobic pocket created by V8 and L12. As depicted in Figure 4.10, the formation of a strong

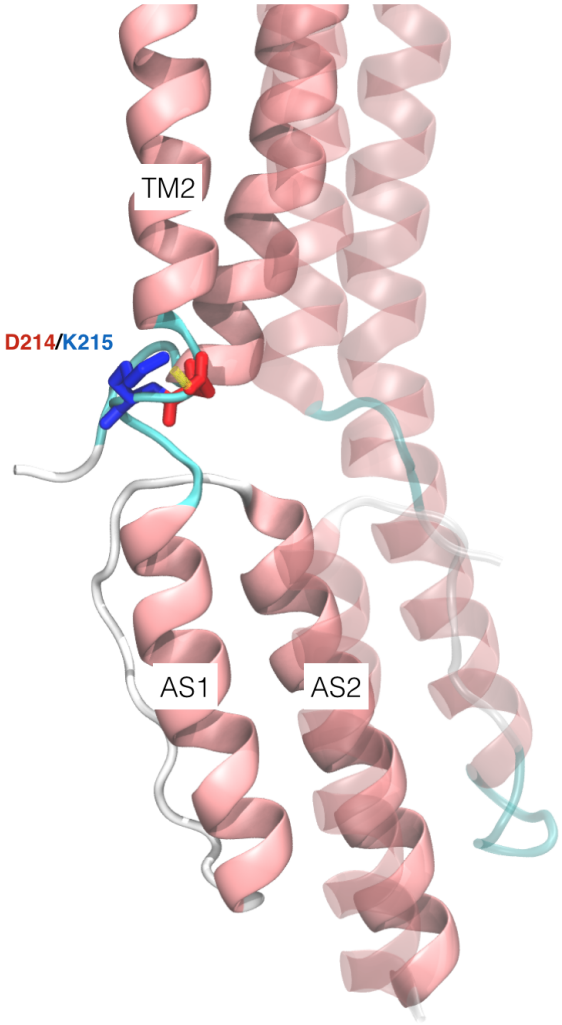

Figure 4.10: I214D mutation reduces helicity of control cable. A salt bridge formed between D214 (red, licorice) and K215 (blue, licorice) disrupts the local backbone hydrogen bonding network, leading the the loss of helical content in the control cable region. Protein colored according to secondary structure: alpha helix (pink), coil (teal), and turn (white).

salt bridge between D214 and K215 in one monomer disrupted the backbone hydrogen bond network of the surrounding area and subsequently led to a loss of helicity in the corresponding control cable.

Hence, our simulations reproduce the key predictions of Parkinson *et al.*, namely that (1) mutations to I214 affect the helicity of the control cable and (2) kinase-ON and kinase-OFF mutations have high and low helical content respectively. In addition, our I214D simulation resolves the mechanism of helical modulation for this particular mutant, namely a salt-bridge-induced disruption of the local hydrogen bonding network. Perhaps proper signaling function may be rescued by removal of the positive charge at residue 215. Additional simulations will be needed to address the affect of I214 mutations on the AS1 helix and HAMP domain as a whole.

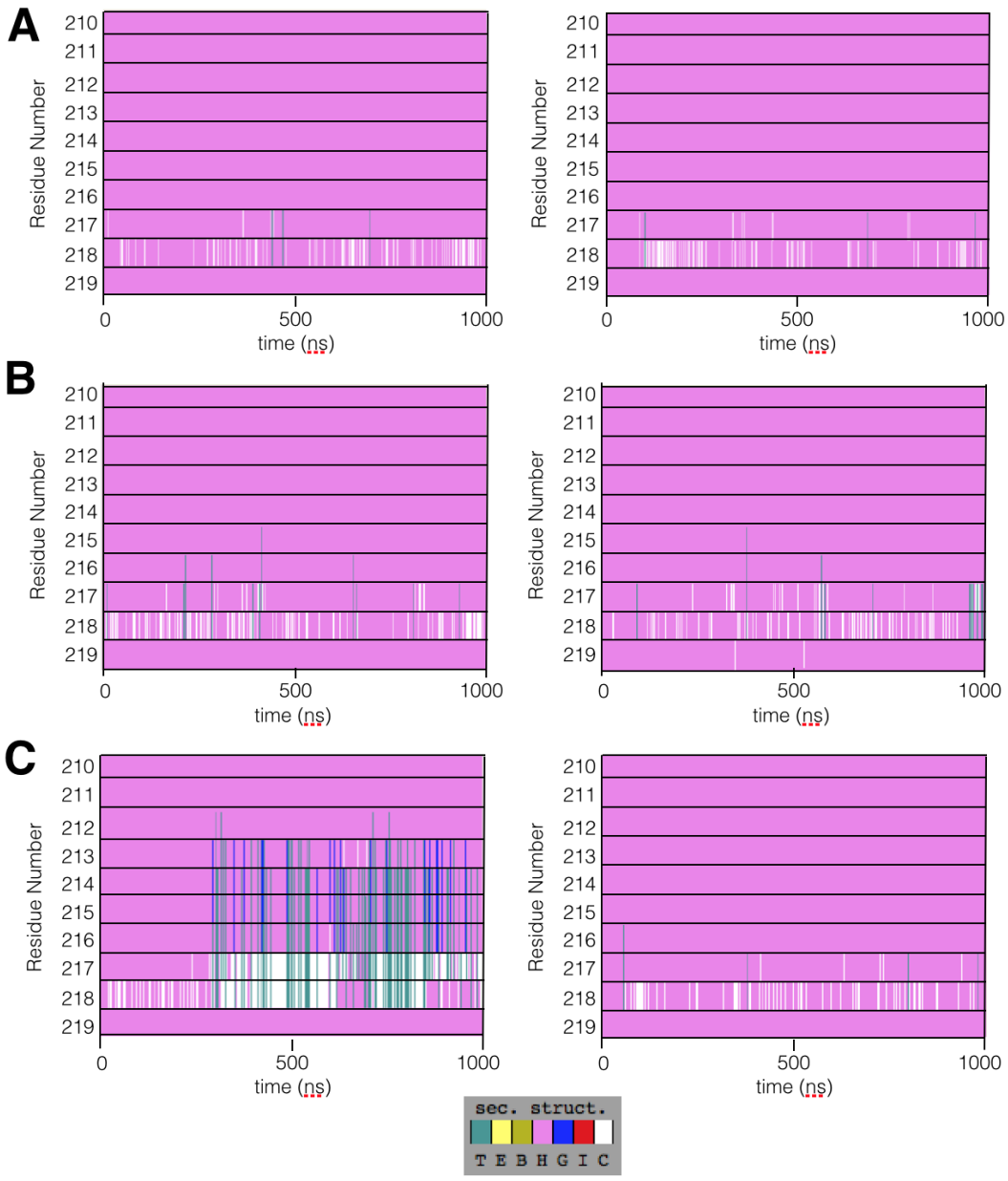

Figure 4.11: Structure of wild-type and mutant control cable regions. Secondary structure of each control cabel from the (A) wild-type, (B) I214A, and (C) I214D Tsr_{TM} systems, plotted over a one microsecond MD simulation. Alpha helices are colored in pink, turns and coils are colored in green and white respectively. In all simulations, L218 is seen to be flexibly undergo reversible helix-coil transitions. In the wild-type and I214A simulations, the control cable (residues 213-217) are seen to form stable alpha helices, whereas the I214D simulation exhibits considerable loss of helicity in this region.

4.2.3 Atomic model and simulations of complete, membrane-bound Tsr homodimer

We then combined our equilibrated Tsr_{TM} model with an X-ray structure of the Tsr cytoplasmic domain (PDB 3ZX6) [124], to construct an atomic model of the complete, membrane-bound *E. coli* Tsr (Fig. 4.12.A). To investigate the dynamical behavior of our intact Tsr model, we then conducted two, 1.5 microsecond, all-atom MD simulations Intriguingly, our simulations identified three areas of increased flexibility as measure by the root-mean-square fluctuation of the aligned backbone atoms. Labeled a, b, and c in Figure 4.12.B, these areas respectively correspond to the (1) serine binding pocket, (2) HAMP domain and methylation bundle, and (3) protein interaction tip. As equilibrium fluctuations are known to play a central role in molecular recognition [147], the flexibility of the binding pocket is not surprising. In addition, the protein interaction tip has been shown computational to be highly dynamic [121]. Interestingly, though, the flexibility of the HAMP and methylation bundle region appears to arise from multiple pivot points (Fig. 4.12.C). Indeed, RMSD-based structural clustering revealed three major classes of "bent" receptor, which exhibited sharp deformations at different points along flexible region b. In particular, bends in the individual classes were seen directly below (i) and above (ii) the methylation bundle as well as directly above the HAMP domain (iii). In the lattermost region, the four-helix TM and HAMP bundles are connected via the two, control cable helices. Hence, the increased flexibility in this region may be attributed to the loss of structural support arising with a tightly packed helical bundle [148]. Bending regions (i) and (ii), however, are situated on either side of the coiled-coil containing the sites of adaptational modification. It is interesting to consider if the flexibility of these regions might be modulated reversible methylation. Finally, the glycine hinge (Fig. 4.12.A, Fig. 4.12.B, label d), a region whose flexibility was previously proposed to play a key role in ON/OFF

switching [116], appears to be relatively rigid, a result also suggested in [121]. It should be noted, though, that the protein interaction tip is unbound (i.e., unconstrained) in our simulations. Hence, within native chemosensory arrays, in which this region is directly bound to CheA-P5, CheW, and two other receptors within a trimer-of-dimers, the receptor cytoplasmic domain is unlikely to have the full range of motions witnessed here. Additional simulations will be required to determine the significance of receptor bending for the transduction of sensory signals.

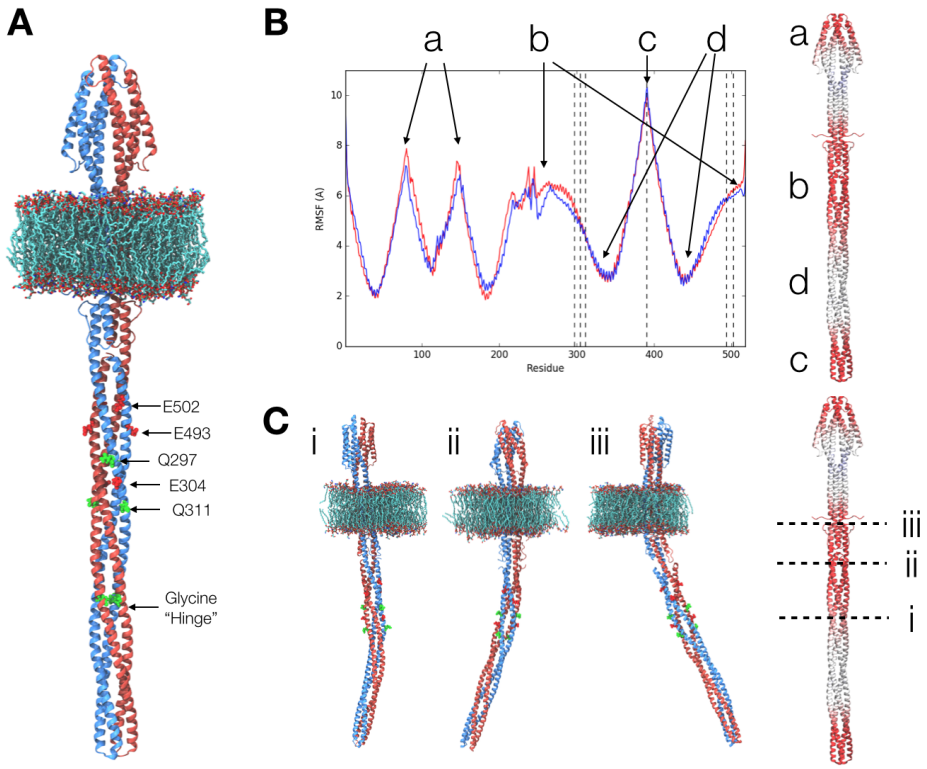

Figure 4.12: (A) Atomic model of membrane-bound E. coli Tsr. The sites of adaptational modification (residues Q297, E304, Q311, E493, and E502) as well as the glycine hinge (residues G340, G341, and G349) are shown in licorice and colored according to their electrostatics: negatively charged (red) and polar (green). Individual monomers are shown in red and blue (B) Root-mean-square fluctuation (RMSF) traces (red and blue) of the backbone atoms of each MCP monomer computed over a one microsecond MD trajectory. RMSF values are mapped onto the Tsr structure using a RWB coloring scheme. Flexible areas of interests are labeled a-d. Methylation sites and the protein interaction tip are marked with a dashed line. (C) Structural clustering separate receptor conformations into several classes (i-iii) with substantial bends near the HAMP and methylation bundles.

4.3 Methods

4.3.1 Molecular dynamics simulations and analysis

All models were hydrated with TIP3P water molecules and subsequently ionized with potassium and chloride ions to a concentration of 150 mM using VMD's [50] solvate and autoionize plugins respectively. Before production runs, each model was subjected to a series of conjugant gradient energy minimizations (30,000 steps in total) and restrained NPT equilibration simulations (10 ns in total). All molecular dynamics simulations were performed using the parallel molecular dynamics code, NAMD 2.10 [49] with the CHARMM36 force field for protein [149] and lipids [150]. Equilibrium simulations were conducted in the NPT ensemble with isobaric and isothermal conditions maintained at 1 atm and 303 K using the Nose-Hoover Langevin piston, with a period 200 femtoseconds (fs) and relaxation time of 50 fs, and the Langevin thermostat with a temperature coupling of 5 ps^{-1}. The r-RESPA integrator scheme [49] with an integration time step of 2 fs was used. SHAKE constraints were applied to all hydrogen atoms [98]. Short-range, non-bonded interactions were calculated every 2 fs with a cutoff of 12 Å and long-range electrostatics were evaluated every 6 fs using the Particle Mesh Ewald (PME) method [99] with a grid size of 1 Å. Periodic boundary conditions were used in all simulations. Visualization and analysis of MD trajectories were performed using VMD [50].

4.4 Supplemental Information

Cysteine(s) in TM1, position(s)	TM1–TM1 cross-linking	Cysteine(s) in TM2, position(s)	TM2–TM2 cross-linking
7	+	189	−
8	−	190	−
9	−	191	−
10	−	192	−
11	++	193	−
10, 12	−	195	−
13	−	196	−
14	+	189, 192, 197	+
15	+	198	−
15, 16	+	200	−
17, 20	+	201, 203	+
18	+	202, 204	−
19	+	203	−
20	−	204	+
21	+	204, 205	+
22	++	206	−
23	−	203, 207, 210	−
24	−	207	−
25	++	208	+
26	−	209	−
27	−	211	+
28	−	210, 212	−
29	++		
30	−		
31	−		
32	−		
33	+		

Figure 4.13: Tar cysteine substitution mutants and their homologous cross-linking efficiencies. +, Detectable cross-linking at 5 to 20% efficiency; ++, >50% efficiency; −, no detectable cross-linking (<5%). Figure reproduced from [130].

Intermolecular disulfide bond, position(s)		Intramolecular disulfide bond, position(s)	
TM1	TM2	TM1	TM2
10/12	208	7	211
10/12	209	10	211
11	210/212	10/12	211
15	208	11	211
19	208	15	204
30	191	21	201/203
30	193	25	189/192/197*
33	189	28	193

Figure 4.14: Tar residue positions of cysteine substitutions that result in efficient (>50%) TM1-TM2 disulfide formation. Residues that could not be unambiguously identified are separated by a /. * This cross-link appears to involve Cys25 and Cys197 because for [Cys^{25},Cys^{189}]- and [Cys^{25},Cys^{192}]-Tar variants heterologous cross-links do not form. Figure reproduced from [130].

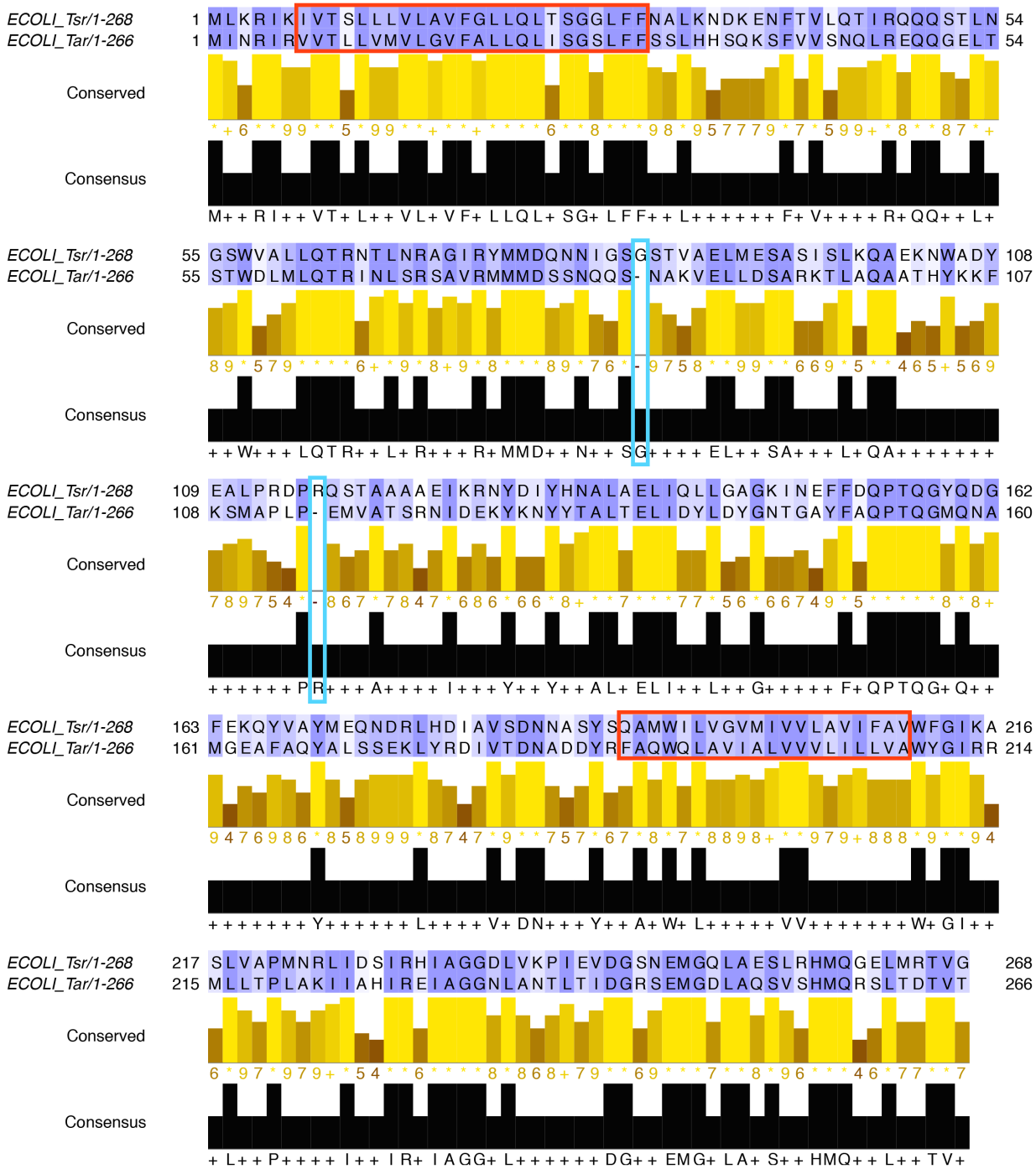

Figure 4.15: MAFFT pairwise sequence alignment between *E. coli* Tsr and Tar periplasmic, TM, and HAMP domains (residues 1-268). Alignment colored using BLOSUM62 score [136]. The conservation of each residue and consensus sequence (showing identical residues) are shown directly below the alignment. Regions of the sequence corresponding to TM1 (residues 7-30) and TM2 (residues 191-210) are highlighted in red. Two gaps (the only two which occur in the alignment of the full Tsr and Tar proteins) are highlighted in blue. Figure created using Jalview [137].

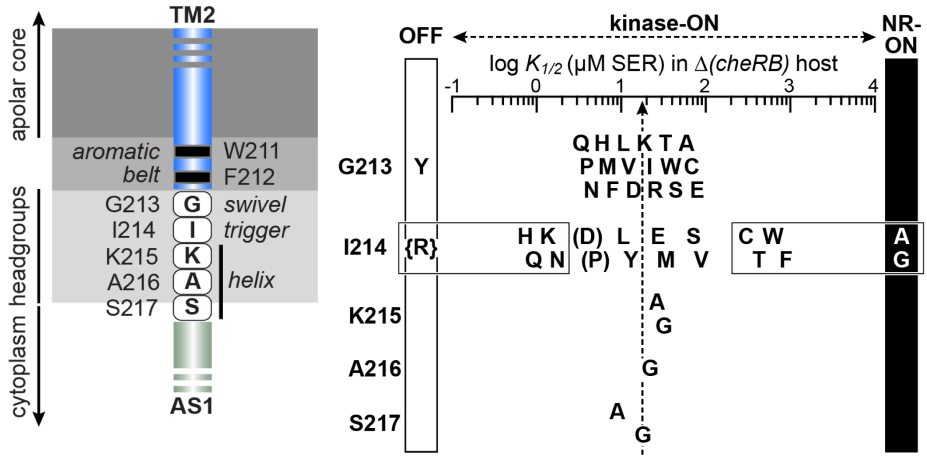

Figure 4.16: Signaling properties of Tsr control cable mutants. (Left) Schematic of TM2, control cable, and HAMP AS1 helix. (Right) Serine response sensitivities of Tsr control cable mutants. Mutations to I214 are seen to give rise to a broad range of signaling phenotypes. I214A and I214D (the two mutants investigated in this study) are seen to be shifted towards the kinase-ON and kinase-OFF state respectively. Figure by Sandy Parkinson.

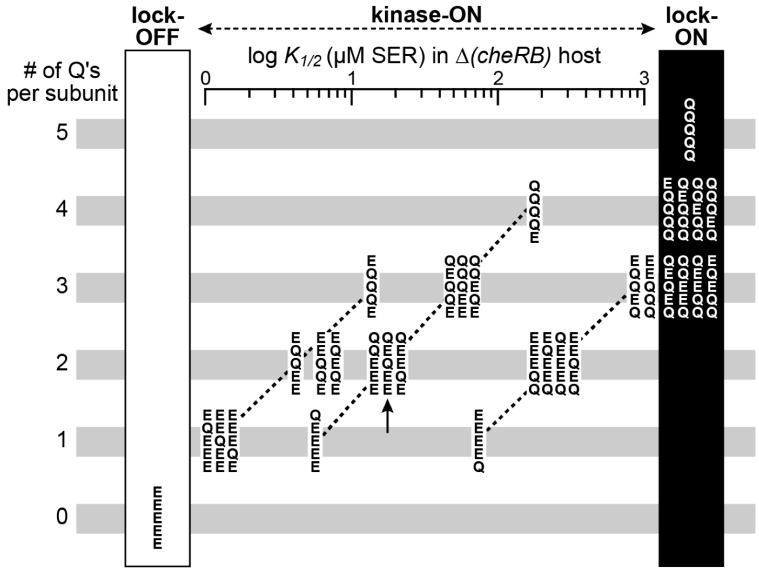

Figure 4.17: Signaling properties of Tsr methylation mutants. Serine response sensitivities of Tsr mutants with varying methylation states and combinatorial orderings at each of its five sites of adaptational modification. Generally speaking, the more methylated the receptor (i.e., the more Q's it has irrespective of ordering), the more its signaling state is shifted towards kinase-ON. Figure by Sandy Parkinson.

Chapter 5

CheY Activation: Effect of Mutation at Key Acetylation Site

Reproduced in part with permission from Milana Fraiberg, Oshri Afanzar, C. Keith Cassidy, Alexandra Gabashvili, Klaus Schulten, Yishai Levin, and Michael Eisenbach. CheY's acetylation sites responsible for generating clockwise flagellar rotation in *Escherichia coli*. *Molecular microbiology*, 95(2), 231-244, 2015.

5.1 Introduction

CheY, depicted in Figure 5.1, is a 128-residue aspartate kinase and the universal response regulator of the two-component bacterial chemotaxis signal transduction system [20, 25, 151]. After the transfer of a phosphate group from activated CheA to its conserved Asp-57 residue, CheY undergoes structural changes that *activate* the protein, enabling it to trigger a cellular response [152, 153]. As discussed in Chapter 1, activated phospho-CheY diffuses to the flagellar motors where it binds to FliM, a protein comprising the motor, to elicit clockwise (CW) rotation and induce tumbling [154, 155]. To return the rotational bias of the motors to their default counter-clockwise (CCW) mode, CheY is rapidly dephosphorylated by CheZ [154, 156].

In addition to phosphorylation, protein acetylation–a reversible, post-translational modification that adds an acetyl group to an amino acid side chain–has been shown to activate CheY both *in vitro* and *in vivo*, enabling it to generate CW rotation [157–160]. This function, termed hereafter "acetate-dependent clockwise generation," is observed even in

the absence of the chemotaxis machinery. Though the sites of acetylation, all lysine (K) residues, have been identified *in vitro*, those affected by acetate *in vivo* and involved in the generation of CW motor rotation have not been. In general, acetylation of lysine (i.e., N-acetylation) is ubiquitous in nature and is important in the regulation of proteins from bacteria to mammals [161,162]

In this study [5], our collaborators in the Eisenbach Lab at the Weizmann Institute used techniques in mass spectrometry to identify K91 and K109 as the major sites responsible for acetate-dependent CW generation *in vivo*. Furthermore, they showed that modification of K91 via acetylation or replacement by specific amino acids increased ability of CheY to generate CW rotation. Hence, non-modified K91 actually represses CW rotation. As shown in Figure 5.4 (Supplemental Information), the K91H mutation was seen to give rise to a especially hyperactive (i.e., unrepressed) phenotype. Generally speaking, because the K91-activating mutations were different from each other in their side chains (including volume and charge), it is likely that their activating effect is produced by the exclusion of lysine from position 91. Hence, to examine how lysine exclusion from position 91 can affect CheY activation at the molecular level, I used all-atom MD simulations for studying the structural and dynamical effects of the K91H mutation, as described in this chapter.

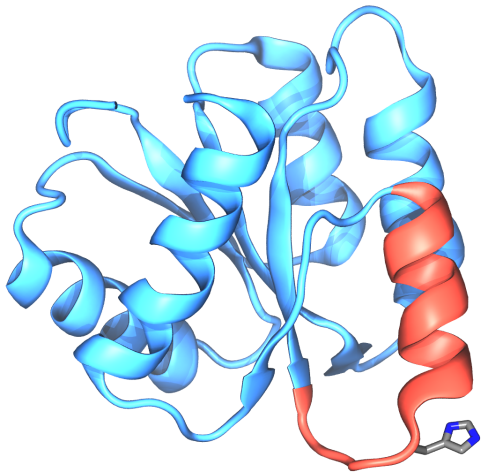

Figure 5.1: Structure of CheY response regulator. CheY, shown in blue, is a 128-residue aspartate kinase and the universal response regulator in bacterial chemotaxis. A key acetylation site, residue 91, is shown mutated to a histidine (licorice), giving rise to enhanced clockwise rotation of the flagellar motors. The $\beta 4\alpha 4$ loop and $\alpha 4$ helix, key areas whose dynamics are affected by the K91H mutation, are shown in red.

5.2 Results

5.2.1 K91H mutation affects conformational dynamics of $\beta 4\alpha 4$ loop

As previous studies had not investigated the effect of the K91H mutation, we chose a comparative computational approach in which we simulated both the mutant of interest and wild-type CheY. Initial atomic coordinates for both the K91H (with histidine protonated at the δ position, Fig. 5.1) and wild-type CheY models were derived from the crystal structure of *E. coli* CheY (PDB 3CHY) [153]. Each system was minimized along with solvent and equilibrated for 250 ns. An ensemble of ten, 250 nanosecond production simulations were subsequently carried out on both the K91H and wild-type systems, resulting in a total of 2.75 microseconds of sampling for each model. Our simulations revealed a higher level of flexibility in the protein regions surrounding position 91 of the K91H system. In particular, residues 88–102, corresponding to the $\beta 4\alpha 4$ loop and $\alpha 4$ helix, showed, on average, an increase in their root mean squared fluctuations (RMSFs) as compared with the corresponding regions in the ensemble of wild-type simulations (Fig. 5.2). Based on this result, we hypothesized that the K91H mutation might give rise to a hyperactive phenotype by altering the conformational dynamics of the $\beta 4\alpha 4$ loop, a region previously shown to correlate with CheY activation [163–165].

5.2.2 Increased flexibility of $\beta 4\alpha 4$ loop favors activated conformation

To characterize the effect of increased flexibility in the protein regions surrounding K91 on the conformational sampling of the $\beta 4\alpha 4$ loop, we used principal components analysis (PCA) [63,166] to provide a concise description of the $\beta 4\alpha 4$ loop conformations witnessed by our MD simulations. The utility of the reduced PCA description to capture the salient features of the dynamics of the $\beta 4\alpha 4$ loop was demonstrated by the relatively small num-

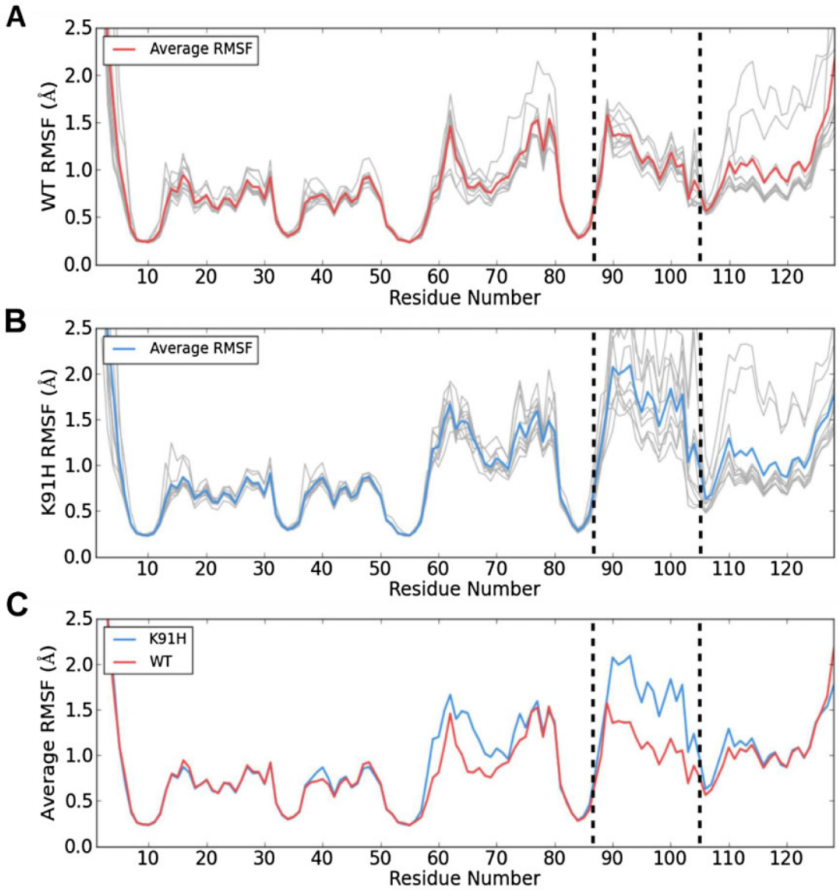

Figure 5.2: K91H exhibits high flexibility of $\beta 4\alpha 4$ loop. RMSF versus residue number for (A) wild-type (WT) and (B) K91H ensemble simulations. Individual RMSF traces for the ten 250 ns simulations in each ensemble (gray lines) were averaged to produce a mean RMSF value for all 129 CheY residues (colored lines). Overlaying the averaged RMSF traces from the wild-type and K91H ensembles (C) highlights regions of differing stiffness. Vertical dotted lines emphasize the increased flexibility in regions of CheY surrounding the K91 mutation (residues 88–105), including the $\beta 4\alpha 4$ loop (residues 88–91).

ber of principal components (PCs) required to describe a significant degree of the conformational variability within the simulation trajectories (Table 5.1, Supplemental Information) [167]. In the present application, the cumulative fractional variance of the top three PCs accounted for over 75% of the total trajectory variance, providing a practical, low-dimensional subspace on to which to project the individual $\beta 4\alpha 4$ loop conformations observed in the K91H and wild-type ensemble simulations. The resulting projections, along with projections of the analogous portions of existing activated CheY structures (PDBs 1FQW, 1ZDM, 1F4V, 1DJM) and the equilibrated, inactive CheY structure, revealed two

well-defined clusters of $\beta 4\alpha 4$ loop conformations (Fig. 5.5, Supplemental Information). Notably, $\beta 4\alpha 4$ loop conformations from both the K91H and wild-type simulations inhabited either cluster, while the active and inactive $\beta 4\alpha 4$ loop conformations localized to single clusters opposite to one another (Fig. 5.5, Supplemental Information). This suggests that both the K91H and wild-type systems sampled distinct functional states corresponding to the active and inactive $\beta 4\alpha 4$ loop conformations in our MD simulations.

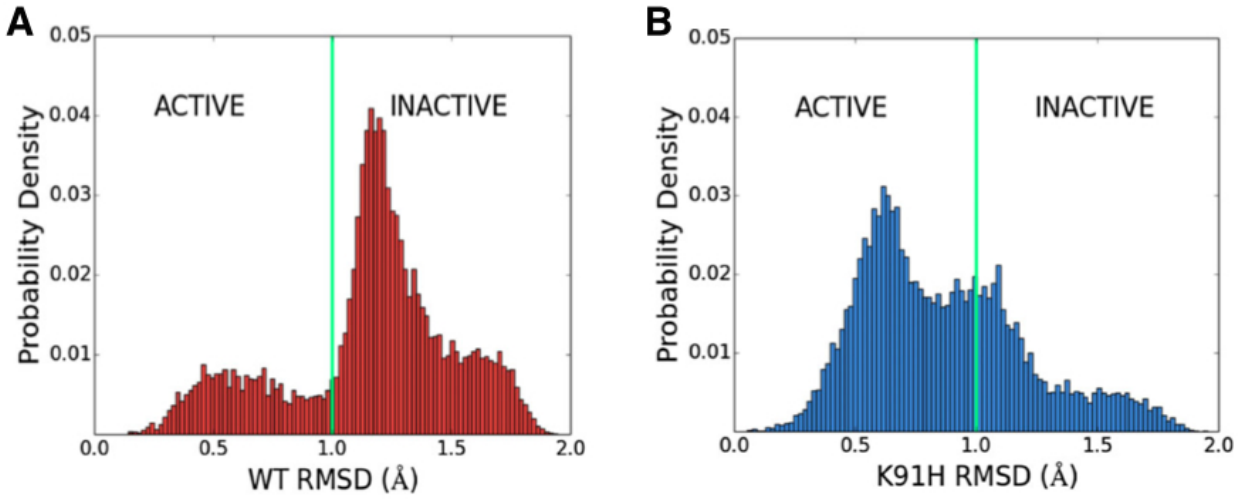

Figure 5.3: K91 represses the active state of CheY's $\beta 4\alpha 4$ loop. Distributions of root mean squared deviations (RMSDs) between projections of an activated $\beta 4\alpha 4$ loop structure (PDB 1FQW) and conformations of the $\beta 4\alpha 4$ loop sampled in the (A) wild-type (WT) and (B) K91H ensemble simulations. Vertical green lines indicate the rough boundary between clusters in principle component space computed using K-means and highlight the population shift incurred by the K91H mutation. See *Simulation analysis* section for more details.

To characterize the organization of projected $\beta 4\alpha 4$ loop conformations in three dimensional (3D) PC space, we used the K-means clustering algorithm [63,168] to systematically assign loop conformations to two clusters (Fig. 5.6, Supplemental Information). Strikingly, K-means revealed that over 80% of the K91H loop conformations were assigned to the same cluster as the activated loop structures, whereas only $\sim$30% of wild-type loop conformations were assigned to this cluster (Table 5.2, Supplemental Information). To further survey the distribution of the conformations in 3D PC space, the root mean squared deviation (RMSD) between projections of the activated $\beta 4\alpha 4$ loop conformations

and those arising within the K91H mutant and wild-type ensemble simulations was calculated. The resulting cumulative RMSD distributions (Fig. 5.3), Supplemental Information) clearly demonstrated an asymmetry in the occupation of the active and inactive $\beta 4\alpha 4$ loop states by the K91H and wild-type CheY systems in agreement with the K-means result (Fig. 5.6. Taken together, the results from our MD simulations and accompanying analysis suggest that the increased flexibility in the $\beta 4\alpha 4$ loop and $\alpha 4$ helix associated with the K91H mutation gives rise to a structural population shift in which the $\beta 4\alpha 4$ loop fluctuates more often near the conformation typical of activated loop structures.

5.3 Discussion

The observations that CW rotation is low whenever K91 is not replaced suggest that K91 represses CW rotation. This repression can be relieved in two ways: acetylation and mutation. MD simulations demonstrated that the $\beta 4\alpha 4$ loop of wild-type CheY co-exists in active and inactive states, spending more time in the inactive state (Fig. 5.3.A). The simulations further demonstrated that when K91 is replaced by histidine, the conformational dynamics of the $\beta 4\alpha 4$ loop are shifted into a flexible regime, leading to the loop spending much more time in the active state (Fig. 5.3.B). Hence, we demonstrate that the molecular mechanism of the K91 CW repression is manifested in the conformational dynamics of the $\beta 4\alpha 4$ loop. Since the K91-activating mutations were different from each other in their side chains and since K91 acetylation is phenotypically similar to K91H replacement, we suggest that the effect of K91 acetylation is mechanistically similar to K91H replacement. Removal of $\beta 4\alpha 4$ loop repression may represent a general activation mechanism in CheY, pertaining also to the canonical phosphorylation activation pathway. Future computational studies will seek to develop parameters for acetylated-lysine and phosphorylated aspartic acid to enable direct simulation of these native activation pathways.

5.4 Supplemental Information

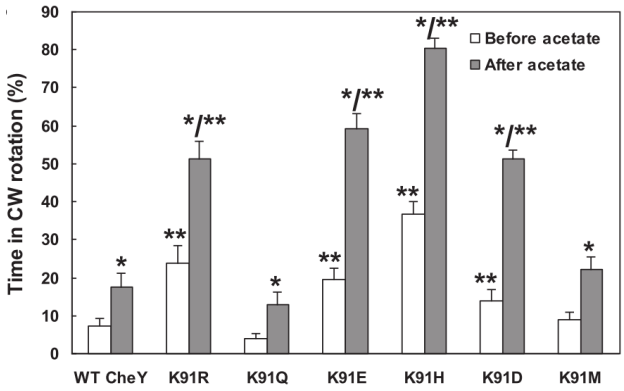

Figure 5.4: Acetate-dependent clockwise generation by CheY containing single replacements at K91. Mutation if K91 to histidine (K91H) ellicits the most hyperactive CW rotation phenotype.

PC #	Fractional Variance (%)	Cumulative Variance (%)
1	42.26	42.26
2	20.22	60.48
3	**13.91**	**76.39**
4	9.03	85.42
5	4.54	89.96
6	2.14	92.10
7	1.94	94.04
8	1.19	95.23
9	1.16	96.39
10	0.72	97.11

Table 5.1: Fractional and cumulative variances of top ten principal components (PCs) obtained from PCA of $\beta 4\alpha 4$ loop conformations.

	% wild-type	% K91H	% Total
Active Cluster (Purple)	28.04	82.66	55.35
Inactive Cluster (Gold)	71.96	17.34	44.65

Table 5.2: Percent of population assigned to two clusters in principal component space (associated with either the active or inactive $\beta 4\alpha 4$ loop) by K-means algorithm.

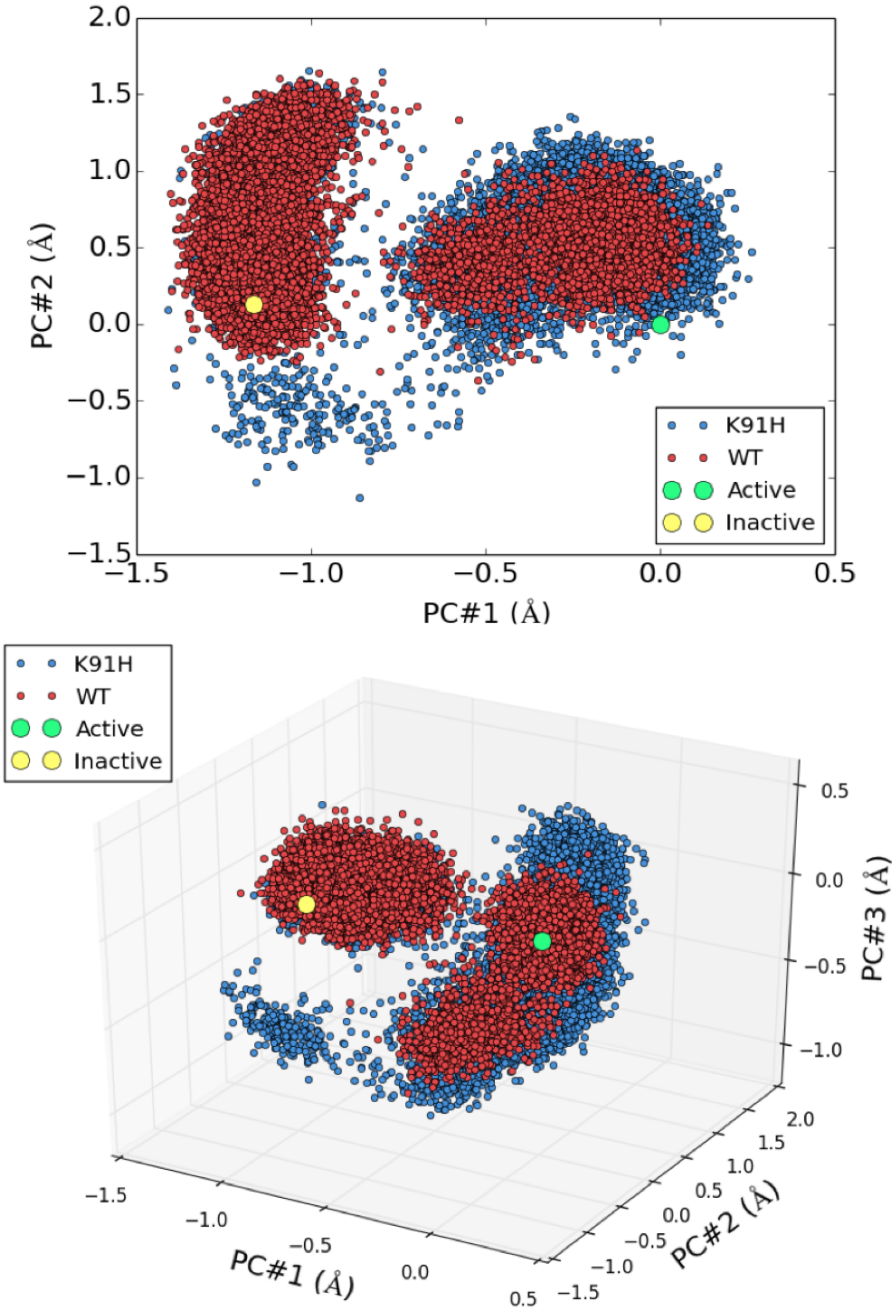

Figure 5.5: Projections of $\beta 4\alpha 4$ loop conformations onto the space spanned by top two (top) and top three (bottom) principal components. 25,000 conformations from the wild-type (red) and K91H (blue) ensemble simulations (representing a total of five microseconds of loop sampling) organize into roughly two clusters. The equilibrated, inactive (yellow) and active (green) $\beta 4\alpha 4$ conformations are seen to associate with distinct populations of sampled conformations.

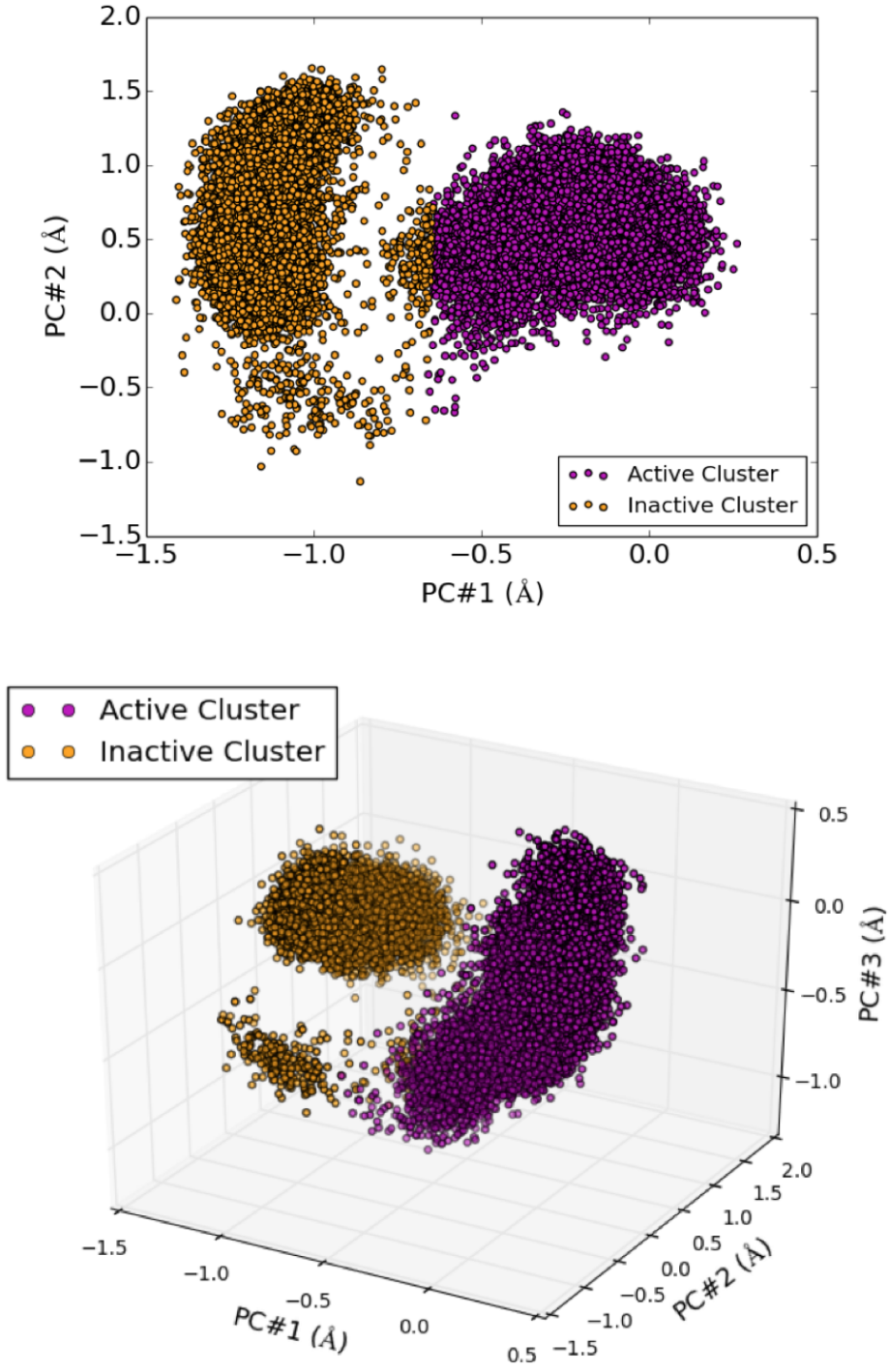

Figure 5.6: Projections of sampled $\beta 4\alpha 4$ loop conformations onto 2D (top) and 3D (bottom) principal component space, colored based on cluster assignment by K-means algorithm (with K = 2 cluster).

5.5 Methods

5.5.1 Molecular dynamics simulations

Atomic coordinates for the initial wild-type and K91H CheY protein models were taken from the crystallographically derived PDB 3CHY [153]. For residues in the PDB in which multiple orientations of the side chains were present, the conformation with the highest reported occupancy was taken. Water molecules present in the PDB were also retained. In the case of the K91H CheY model, lysine 91 was replaced with a histidine protonated at the δ position using the psfgen module in VMD [50]. Models were then hydrated with TIP3P water molecules using VMD's solvate plugin, resulting in an orthorhombic simulation box of size 64x59x60 Å. Using VMD's autoionize plugin, the wild-type and K91H systems were then neutralized by adding four and five sodium ions, respectively, and subsequently ionized with sodium and chloride ions to the physiological concentration of 150 mM. The resulting wild-type and K91H systems contained 21,191 and 21,193 atoms, respectively, including protein, water and ions.

All MD simulations were performed using the parallel MD code, NAMD 2.9 [49], and CHARMM22 force field [96] with CMAP corrections [97]. Simulations were conducted in the NPT ensemble with isobaric and isothermal conditions maintained at 1 atm and 310 K using the Nose-Hoover Langevin piston with period 200 fs and decay 50 fs and the Langevin thermostat with a temperature coupling of 5 ps^{-1}. The r-RESPA integrator scheme [49] with an integration time step of 2 fs was used in all simulations. SHAKE constraints were applied to all hydrogen atoms [98]. Short-range, non-bonded interactions, calculated every 2 fs with a cutoff of 12 Å and long-range electrostatics, were evaluated every 6 fs using the Particle Mesh Ewald method [99] with a grid size of 1 Å. Periodic boundary conditions were used in all simulations. Both wild-type and K91H models were subjected to a series of conjugant gradient energy minimizations: 100,000 steps with

all protein atoms excluding hydrogen restrained, 100,000 steps with backbone atoms restrained and 100,000 steps with all atoms unrestrained. Before proceeding with unrestrained equilibration simulations, each model underwent a 2 ns NPT simulation in which backbone atoms were harmonically restrained with spring constant = 0.5 kcal/mol*nm^2.

5.5.2 Simulation analysis

Visualization and extraction of raw trajectory data for analysis were performed using VMD [50]. RMSF calculations were performed using the measure function in VMD. Clustering analysis was performed using the implementation of the K-means algorithm from the python scientific computing library, Scipy. PCA was performed using the protein dynamics analysis package, ProDy [169]. Before conducting PCA, the backbone atomic coordinates of the $\beta 4\alpha 4$ loop (residues 88–91) were extracted from the concatenated K91H and wild-type ensemble simulations and aligned using root mean square deviation fitting in VMD. A total of 25,000 conformations (12,500 per model), sampled over 5 µs of simulation, were analyzed. The top three resulting PCs were used as a basis on which the backbone structures of the $\beta 4\alpha 4$ loop were projected for further analysis. The RMSD values depicted in Figure 5.6 were calculated in the three-dimensional PC space described above and are equivalent to the Euclidean distances between projections of a representative activated loop conformation (PDB 1FQW) and those arising in the K91H and wild-type ensemble simulation trajectories. Illustrations of K-means and PCA results were produced using the python plotting library, Matplotlib [101].

5.5.3 Experimental methods

A discussion concerning the experimental methodologies used in this study can be found in [5].

Chapter 6

Outlook: The *E. coli* Transmembrane Chemosensory Array

In this dissertation, I have presented work seeking to provide a much-needed molecular platform from which to rationally design investigations of chemosensory array structure and function. Towards this end, I have attempted to use computational modeling and molecular dynamics simulations to synthesize existing and cutting-edge structural data into high-fidelity atomic models. In addition to the work presented here, I have recently developed atomic homology models of the previously uncharacterized *E. coli* chemotaxis signaling proteins and core signaling complexes. Moreover, using the lipid-monolayer system described in Chapter 3, members of Peijun Zhang's Lab have since resolved density maps of the array in multiple, well-defined signaling states and further improved the resolution of their data to $\sim$1 nm. Taken together, this progress has enabled the construction of the first atomic model of the intact, *E. coli* transmembrane chemosensory array (Fig. 6.1). In the coming months, I will assemble from Zhang *et al.*'s multi-state data a series of such models, enabling an atomistically resolved analysis of signaling-related structural events. With hope, the resulting models and simulations will provide directly transferrable structural and dynamical predictions, facilitating the unification of the wealth of existing biochemical and biophysical data and helping to elucidate a complete mechanistic description of signal transduction and amplification within this truly impressive biological sensory apparatus.

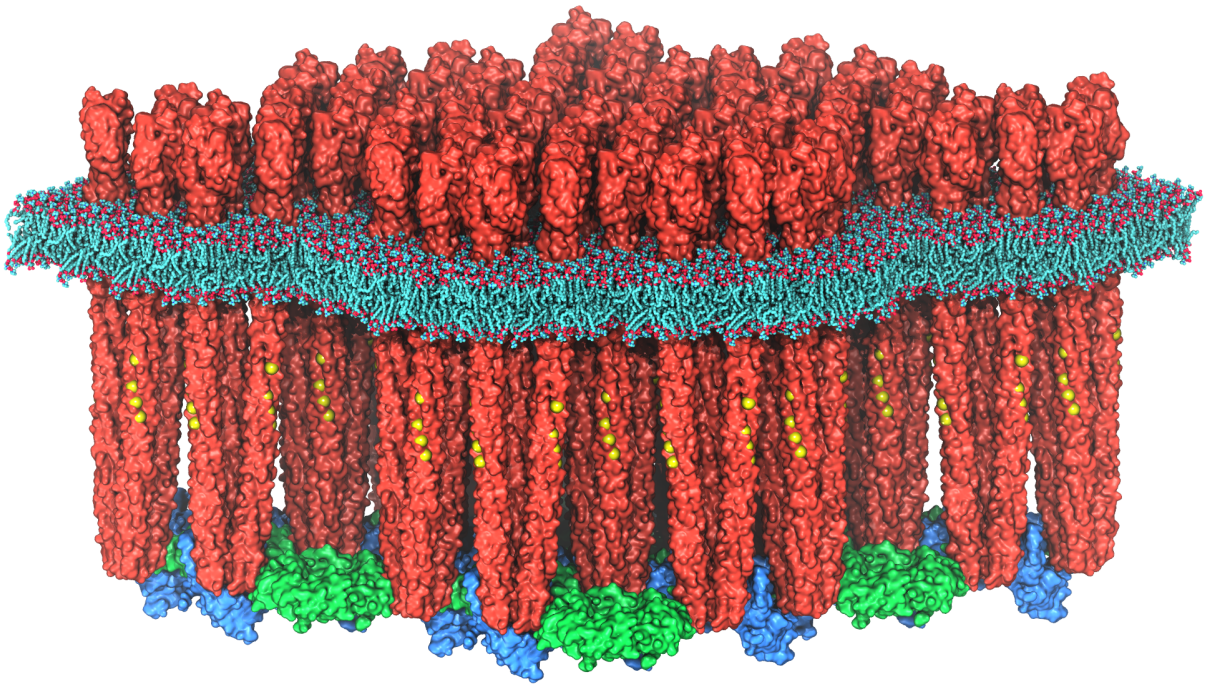

Figure 6.1: The *E. coli* transmembrane chemosensory array

Bibliography

[1] J. E. Stone, J. R. Perilla, C. K. Cassidy, and K. Schulten, "Gpu-accelerated molecular dynamics clustering analysis with openacc," in Parallel Programming with OpenACC (R. Farber, ed.), ch. 11, pp. 215–240, Cambridge, MA: Elsevier, 2016.

[2] C. K. Cassidy, B. A. Himes, F. J. Alvarez, J. Ma, G. Zhao, J. R. Perilla, K. Schulten, and P. Zhang, "CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling," eLife, vol. 4, p. e08419, 2015.

[3] B. C. Goh, J. A. Hadden, R. C. Bernardi, A. Singharoy, R. McGreevy, T. Rudack, C. K. Cassidy, and K. Schulten, "Computational methodologies for real-space structural refinement of large macromolecular complexes," Annual review of biophysics, no. 0, 2016.

[4] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, "Molecular dynamics simulations of large macromolecular complexes," Current opinion in structural biology, vol. 31, pp. 64–74, 2015.

[5] M. Fraiberg, O. Afanzar, C. K. Cassidy, A. Gabashvili, K. Schulten, Y. Levin, and M. Eisenbach, "CheY's acetylation sites responsible for generating clockwise flagellar rotation in *Escherichia coli*," Molecular microbiology, vol. 95, pp. 231–244, 2015.

[6] M. Eisenbach, Chemotaxis. London: Imperial College Press, 2004.

[7] U. B. Kaupp, N. D. Kashikar, and I. Weyand, "Mechanisms of sperm chemotaxis," Annu. Rev. Physiol., vol. 70, pp. 93–117, 2008.

[8] J. Franca-Koh, Y. Kamimura, and P. Devreotes, "Navigating signaling networks: chemotaxis in *Dictyostelium discoideum*," Current opinion in genetics & development, vol. 16, no. 4, pp. 333–338, 2006.

[9] E. Kolaczkowska and P. Kubes, "Neutrophil recruitment and function in health and inflammation," Nature Reviews Immunology, vol. 13, no. 3, pp. 159–175, 2013.

[10] P. N. Devreotes and S. H. Zigmond, "Chemotaxis in eukaryotic cells: a focus on leukocytes and Dictyostelium," Annual review of cell biology, vol. 4, no. 1, pp. 649–686, 1988.

[11] G. H. Wadhams and J. P. Armitage, "Making sense of it all: bacterial chemotaxis," Nature Reviews Molecular Cell Biology, vol. 5, no. 12, pp. 1024–1037, 2004.

[12] E. Aihara, C. Closson, A. L. Matthis, M. A. Schumacher, A. C. Engevik, Y. Zavros, K. M. Ottemann, and M. H. Montrose, "Motility and chemotaxis mediate the preferential colonization of gastric injury sites by *Helicobacter pylori*," PLoS Pathog, vol. 10, no. 7, p. e1004275, 2014.

[13] R. Lux, A. Moter, and W. Shi, "Chemotaxis in pathogenic spirochetes: directed movement toward targeting tissues?," Journal of molecular microbiology and biotechnology, vol. 2, no. 4, pp. 355–364, 2000.

[14] R. Lux and W. Shi, "Chemotaxis-guided movements in bacteria," Critical Reviews in Oral Biology & Medicine, vol. 15, no. 4, pp. 207–220, 2004.

[15] K. M. Ottemann and J. F. Miller, "Roles for motility in bacterial–host interactions," Molecular microbiology, vol. 24, no. 6, pp. 1109–1117, 1997.

[16] J. Adler, "Chemotaxis in bacteria," Science, vol. 153, no. 3737, pp. 708–716, 1966.

[17] J. Adler, G. L. Hazelbauer, and M. M. Dahl, "Chemotaxis toward sugars in *Escherichia coli*," Journal of Bacteriology, vol. 115, no. 3, pp. 824–847, 1973.

[18] H. C. Berg, E. coli in Motion. Springer, 2004.

[19] E. J. Capra and M. T. Laub, "Evolution of two-component signal transduction systems.," Annual review of microbiology, vol. 66, pp. 325–47, 2012.

[20] J. J. Falke, R. B. Bass, S. L. Butler, S. A. Chervitz, and M. A. Danielson, "The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes," Annual review of cell and developmental biology, vol. 13, p. 457, 1997.

[21] P. Casino, V. Rubio, and A. Marina, "The mechanism of signal transduction by two-component systems.," Current opinion in structural biology, vol. 20, no. 6, pp. 763–71, 2010.

[22] H. C. Berg, "Chemotaxis in bacteria.," Annual review of biophysics and bioengineering, vol. 4, no. 00, pp. 119–36, 1975.

[23] G. L. Hazelbauer, "Bacterial chemotaxis: the early years of molecular studies," Annual review of microbiology, vol. 66, p. 285, 2012.

[24] J. B. Stock, G. S. Lukat, and A. M. Stock, "Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks," Annual review of biophysics and biophysical chemistry, vol. 20, no. 1, pp. 109–136, 1991.

[25] H. Szurmant and G. Ordal, "Diversity in chemotaxis mechanisms among the bacteria and archaea," Microbiology and Molecular Biology, vol. 68, no. 2, 2004.

[26] S. Djordjevic and A. M. Stock, "Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding S-adenosylmethionine," Structure, vol. 5, no. 4, pp. 545–558, 1997.

[27] A. H. West, E. Martinez-Hackert, and A. M. Stock, "Crystal structure of the catalytic domain of the chemotaxis receptor methylesterase, CheB," Journal of molecular biology, vol. 250, no. 2, pp. 276–290, 1995.

[28] R. M. Weis and D. E. Koshland, "Reversible receptor methylation is essential for normal chemotaxis of *Escherichia coli* in gradients of aspartic acid.," Proceedings of the National Academy of Sciences, vol. 85, no. 1, pp. 83–7, 1988.

[29] M. F. Goy, M. S. Springer, and J. Adler, "Sensory transduction in *Escherichia coli*: role of a protein methylation reaction in sensory adaptation," Proceedings of the National Academy of Sciences, vol. 74, no. 11, pp. 4964–4968, 1977.

[30] G. L. Hazelbauer and W.-C. Lai, "Bacterial chemoreceptors: providing enhanced features to two-component signaling," Current opinion in microbiology, vol. 13, no. 2, pp. 124–132, 2010.

[31] J. S. Parkinson, G. L. Hazelbauer, and J. J. Falke, "Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update," Trends in Microbiology, pp. 1–10, 2015.

[32] J. J. Falke and K. N. Piasta, "Architecture and signal transduction mechanism of the bacterial chemosensory array: progress, controversies, and challenges," Current opinion in structural biology, vol. 29, pp. 85–94, 2014.

[33] A. Briegel, D. R. Ortega, E. I. Tocheva, K. Wuichet, Z. Li, S. Chen, A. Müller, C. V. Iancu, G. E. Murphy, M. J. Dobro, I. B. Zhulin, and G. J. Jensen, "Universal architecture of bacterial chemoreceptor arrays," Proceedings of the National Academy of Sciences, vol. 106, no. 40, pp. 17181–17186, 2009.

[34] A. Briegel, M. L. Wong, H. L. Hodges, C. M. Oikonomou, K. N. Piasta, M. J. Harris, D. J. Fowler, L. K. Thompson, J. J. Falke, L. L. Kiessling, and G. J. Jensen, "New insights into bacterial chemoreceptor array structure and assembly from electron cryotomography.," Biochemistry, 2014.

[35] A. Briegel, D. R. Ortega, A. N. Huang, C. M. Oikonomou, R. P. Gunsalus, and G. J. Jensen, "Structural conservation of chemotaxis machinery across archaea and bacteria," Environmental Microbiology Reports, vol. 7, no. 3, pp. 414–419, 2015.

[36] M. Li and G. L. Hazelbauer, "Core unit of chemotaxis signaling complexes.," Proceedings of the National Academy of Sciences, vol. 108, no. 23, pp. 9390–5, 2011.

[37] A. Briegel, M. S. Ladinsky, C. Oikonomou, C. W. Jones, M. J. Harris, D. J. Fowler, Y.-W. Chang, L. K. Thompson, J. P. Armitage, and G. J. Jensen, "Structure of bacterial cytoplasmic chemoreceptor arrays and implications for chemotactic signaling," eLife, vol. 3, 2014.

[38] A. Briegel, X. Li, A. M. Bilwes, K. T. Hughes, G. J. Jensen, and B. R. Crane, "Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins.," Proceedings of the National Academy of Sciences, vol. 109, no. 10, pp. 3766–71, 2012.

[39] J. Liu, B. Hu, D. R. Morado, S. Jani, M. D. Manson, and W. Margolin, "Molecular architecture of chemoreceptor arrays revealed by cryoelectron tomography of *Escherichia coli* minicells.," Proceedings of the National Academy of Sciences, vol. 109, no. 23, pp. E1481–8, 2012.

[40] G. L. Hazelbauer, J. J. Falke, and J. S. Parkinson, "Bacterial chemoreceptors: high-performance signaling in networked arrays.," Trends in biochemical sciences, vol. 33, no. 1, pp. 9–19, 2008.

[41] V. Sourjik and N. S. Wingreen, "Responding to chemical gradients: bacterial chemotaxis.," Current opinion in cell biology, vol. 24, no. 2, pp. 262–8, 2012.

[42] Y. Tu, "Quantitative modeling of bacterial chemotaxis: signal amplification and accurate adaptation," Annual review of biophysics, vol. 42, p. 337, 2013.

[43] D. Bray, M. Levin, and C. Morton-Firth, "Receptor clustering as a cellular mechanism to control sensitivity," Nature, vol. 393, 1998.

[44] R. Mesibov, G. W. Ordal, and J. Adler, "The range of attractant concentrations for bacterial chemotaxis and the threshold and size of response over this range. Weber law and related phenomena.," The Journal of general physiology, vol. 62, no. 2, pp. 203–223, 1973.

[45] N. Barkal and S. Leibler, "Robustness in simple biochemical networks," Nature, pp. 913–917, 1997.

[46] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw, "Biomolecular simulation: A computational microscope for molecular biology," Annual review of biophysics, vol. 41, no. 1, pp. 429–452, 2012.

[47] M. P. Allen and D. J. Tildesley, Computer Simulation of Liquids. New York: Oxford University Press, 1987.

[48] D. Frenkel and B. Smit, Understanding Molecular Simulation From Algorithms to Applications. California: Academic Press, 2002.

[49] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with NAMD," Journal of computational chemistry, vol. 26, no. 16, pp. 1781–1802, 2005.

[50] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," Journal of molecular graphics, vol. 14, no. 1, pp. 33–38, 1996.

[51] L. G. Trabuco, E. Villa, K. Mitra, J. Frank, and K. Schulten, "Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics," Structure, vol. 16, no. 5, pp. 673–683, 2008.

[52] R. McGreevy, I. Teo, A. Singharoy, and K. Schulten, "Advances in the molecular dynamics flexible fitting method for cryo-EM modeling," Methods, vol. 100, pp. 50–60, 2016.

[53] L. G. Trabuco, E. Villa, E. Schreiner, C. B. Harrison, and K. Schulten, "Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and x-ray crystallography," Methods, vol. 49, no. 2, pp. 174–180, 2009.

[54] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang, "Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics," Nature, vol. 497, pp. 643–646, 2013.

[55] A. Rajan, P. L. Freddolino, and K. Schulten, "Going beyond clustering in md trajectory analysis: an application to villin headpiece folding," PLoS One, vol. 5, no. 4, p. e9890, 2010.

[56] P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, "Challenges in protein-folding simulations," Nature physics, vol. 6, no. 10, pp. 751–758, 2010.

[57] B. C. Goh, J. R. Perilla, M. R. England, K. J. Heyrana, R. C. Craven, and K. Schulten, "Atomic modeling of an immature retroviral lattice using molecular dynamics and mutagenesis," Structure, vol. 23, no. 8, pp. 1414–1425, 2015.

[58] A. P. Reynolds, G. Richards, B. de la Iglesia, and V. J. Rayward-Smith, "Clustering rules: a comparison of partitioning and hierarchical clustering algorithms," Journal of Mathematical Modelling and Algorithms, vol. 5, no. 4, pp. 475–504, 2006.

[59] P. L. Freddolino and K. Schulten, "Common structural transitions in explicit-solvent simulations of villin headpiece folding," Biophysical journal, vol. 97, pp. 2338–2347, 2009.

[60] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, vol. 32, no. 5, pp. 922–923, 1976.

[61] W. Kabsch, "A discussion of the solution for the best rotation to relate two sets of vectors," Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, vol. 34, no. 5, pp. 827–828, 1978.

[62] I. Jolliffe, Principal Component Analysis. New York: Springer, 2nd ed., 2002.

[63] C. M. Bishop, "Pattern recognition," Machine Learning, vol. 128, 2006.

[64] M. Ringnér, "What is principal component analysis?," Nature biotechnology, vol. 26, no. 3, pp. 303–304, 2008.

[65] V. Sourjik and J. P. Armitage, "Spatial organization in bacterial chemotaxis," The EMBO journal, vol. 29, no. 16, pp. 2724–2733, 2010.

[66] V. Sourjik and H. Berg, "Receptor sensitivity in bacterial chemotaxis," Proceedings of the National Academy of Sciences, vol. 99, no. 1, 2002.

[67] F. W. Dahlquist, "Amplification of signaling events in bacteria," Science Signaling, vol. 2002, no. 132, p. pe24, 2002.

[68] G. H. Wadhams and J. P. Armitage, "Making sense of it all: bacterial chemotaxis," Nature Reviews Molecular Cell Biology, vol. 5, no. 12, pp. 1024–1037, 2004.

[69] K. Kim, H. Yokota, and S. Kim, "Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor," Nature, vol. 400, no. August, pp. 787–792, 1999.

[70] A. Bilwes, L. Alex, B. Crane, and M. Simon, "Structure of CheA, a signal-transducing histidine kinase," Cell, vol. 96, pp. 131–141, 1999.

[71] S.-Y. Park, P. P. Borbat, G. Gonzalez-Bonet, J. Bhatnagar, A. M. Pollard, J. H. Freed, A. M. Bilwes, and B. R. Crane, "Reconstruction of the chemotaxis receptor–kinase assembly," Nature structural & molecular biology, vol. 13, no. 5, pp. 400–407, 2006.

[72] Y. Li, Y. Hu, W. Fu, B. Xia, and C. Jin, "Solution structure of the bacterial chemotaxis adaptor protein CheW from *Escherichia coli*," Biochemical and biophysical research communications, vol. 360, no. 4, pp. 863–867, 2007.

[73] I. J. Griswold, H. Zhou, M. Matison, R. V. Swanson, L. P. McIntosh, M. I. Simon, and F. W. Dahlquist, "The solution structure and interactions of chew from *Thermotoga maritima*," Nature Structural & Molecular Biology, vol. 9, no. 2, pp. 121–125, 2002.

[74] X. Li, A. D. Fleetwood, C. Bayas, A. M. Bilwes, D. R. Ortega, J. J. Falke, I. B. Zhulin, and B. R. Crane, "The 3.2 Å resolution structure of a receptor: CheA:CheW signaling complex defines overlapping binding sites and key residue interactions within bacterial chemosensory arrays," Biochemistry, vol. 52, no. 22, pp. 3852–3865, 2013.

[75] J. Bhatnagar, P. P. Borbat, A. M. Pollard, A. M. Bilwes, J. H. Freed, and B. R. Crane, "Structure of the ternary complex formed by a chemotaxis receptor signaling domain, the CheA histidine kinase, and the coupling protein CheW as determined by pulsed dipolar ESR spectroscopy.," Biochemistry, vol. 49, no. 18, pp. 3824–41, 2010.

[76] A. Vu, X. Wang, H. Zhou, and F. W. Dahlquist, "The receptor—CheW binding interface in bacterial chemotaxis," Journal of molecular biology, vol. 415, no. 4, pp. 759–767, 2012.

[77] X. Wang, A. Vu, K. Lee, and F. W. Dahlquist, "CheA—receptor interaction sites in bacterial chemotaxis," Journal of molecular biology, vol. 422, no. 2, pp. 282–290, 2012.

[78] M. Li, C. M. Khursigara, S. Subramaniam, and G. L. Hazelbauer, "Chemotaxis kinase CheA is activated by three neighbouring chemoreceptor dimers as effectively as by receptor clusters.," Molecular microbiology, vol. 79, no. 3, pp. 677–85, 2011.

[79] K. N. Piasta, C. J. Ulliman, P. F. Slivka, B. R. Crane, and J. J. Falke, "Defining a key receptor-CheA kinase contact and elucidating its function in the membrane-bound bacterial chemosensory array: A disulfide mapping and TAM-IDS study," Biochemistry, vol. 52, no. 22, pp. 3866–3880, 2013.

[80] A. M. Natale, J. L. Duplantis, K. N. Piasta, and J. J. Falke, "Structure, function, and on-off switching of a core unit contact between CheA kinase and CheW adaptor protein in the bacterial chemosensory array: A disulfide mapping and mutagenesis study.," Biochemistry, vol. 52, no. 44, pp. 7753–65, 2013.

[81] M. Li and G. L. Hazelbauer, "Selective allosteric coupling in core chemotaxis signaling complexes.," Proceedings of the National Academy of Sciences, vol. 2014, no. 6, pp. 1–6, 2014.

[82] P. Zhang, C. M. Khursigara, L. M. Hartnell, and S. Subramaniam, "Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy.," Proceedings of the National Academy of Sciences, vol. 104, no. 10, pp. 3777–81, 2007.

[83] A. Briegel, P. Ames, J. C. Gumbart, C. M. Oikonomou, J. S. Parkinson, and G. J. Jensen, "The mobility of two kinase domains in the *Escherichia coli* chemoreceptor array varies with signalling state.," Molecular microbiology, vol. 89, no. 5, pp. 831–41, 2013.

[84] D. J. Montefusco, A. L. Shrout, T. Y. Besschetnova, and R. M. Weis, "Formation and activity of template-assembled receptor signaling complexes.," Langmuir : the ACS journal of surfaces and colloids, vol. 23, no. 6, pp. 3280–3289, 2007.

[85] A. Shrout, D. Montefusco, and R. Weis, "Template-directed assembly of receptor signaling complexes," Biochemistry, pp. 13379–13385, 2003.

[86] D. W. Taylor, D. F. Kelly, A. Cheng, and K. A. Taylor, "On the freezing and identification of lipid monolayer 2-D arrays for cryoelectron microscopy," Journal of Structural Biology, vol. 160, no. 3, pp. 305–312, 2007.

[87] K. A. Taylor and D. W. Taylor, "Structural studies of cytoskeletal protein arrays formed on lipid monolayers," Journal of Structural Biology, vol. 128, no. 1, pp. 75–81, 1999.

[88] J. J. Fernández, S. Li, and R. A. Crowther, "CTF determination and correction in electron cryotomography," Ultramicroscopy, vol. 106, no. 7, pp. 587–596, 2006.

[89] W. O. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," Journal of Microscopy, vol. 127, no. 2, pp. 127–138, 1982.

[90] D. R. Ortega, G. Mo, K. Lee, H. Zhou, J. Baudry, F. W. Dahlquist, and I. B. Zhulin, "Conformational coupling between receptor and kinase binding sites through a conserved salt bridge in a signaling complex scaffold protein.," PLoS computational biology, vol. 9, no. 11, p. e1003337, 2013.

[91] K. N. Piasta and J. J. Falke, "Increasing and decreasing the ultrastability of bacterial chemotaxis core signaling complexes by modifying protein-protein contacts," Biochemistry, vol. 53, no. 35, pp. 5592–5600, 2014.

[92] A. H. Erbse and J. J. Falke, "The core signaling proteins of bacterial chemotaxis assemble to form an ultrastable complex," Biochemistry, vol. 48, no. 29, pp. 6975–6987, 2009.

[93] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," Nucleic acids research, vol. 28, no. 1, pp. 235–242, 2000.

[94] X. Wang, P. Vallurupalli, A. Vu, K. Lee, S. Sun, W. J. Bai, C. Wu, H. Zhou, J. E. Shea, L. E. Kay, and F. W. Dahlquist, "The linker between the dimerization and catalytic domains of the CheA histidine kinase propagates changes in structure and dynamics that are important for enzymatic activity," Biochemistry, vol. 53, no. 5, pp. 855–861, 2014.

[95] D. Müllner, "fastcluster: Fast hierarchical, agglomerative clustering routines for R and python," Journal of Statistical Software, vol. 53, no. 9, pp. 1–18, 2013.

[96] B. R. Brooks, C. L. Brooks, A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, and et al., "CHARMM: the biomolecular simulation program," Journal of computational chemistry, vol. 30, no. 10, pp. 1545–1614, 2009.

[97] A. D. MacKerell Jr, M. Feig, and C. L. Brooks, "Improved treatment of the protein backbone in empirical force fields," Journal of the American Chemical Society, vol. 126, no. 3, pp. 698–699, 2003.

[98] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," Journal of Computational Physics, vol. 23, no. 3, pp. 327–341, 1977.

[99] T. Darden, D. York, and L. Pedersen, "Particle Mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems," The Journal of chemical physics, vol. 98, pp. 10089–10092, 1993.

[100] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," Journal of computational chemistry, vol. 32, no. 10, pp. 2319–2327, 2011.

[101] J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing in science and engineering, vol. 9, no. 3, pp. 90–95, 2007.

[102] J. B. Stock, M. N. Levit, and P. M. Wolanin, "Information processing in bacterial chemotaxis," Science Signaling, vol. 2002, no. 132, pp. pe25–pe25, 2002.

[103] R. B. Bourret and A. M. Stock, "Molecular information processing: lessons from bacterial chemotaxis," Journal of Biological Chemistry, vol. 277, no. 12, pp. 9625–9628, 2002.

[104] P. M. Wolanin and J. B. Stock, "Bacterial chemosensing: cooperative molecular logic," Current biology, vol. 14, no. 12, pp. R486–R487, 2004.

[105] K. M. Ottemann, "A piston model for transmembrane signaling of the aspartate receptor," Science, vol. 285, no. 5434, pp. 1751–1754, 1999.

[106] J. J. Falke and G. L. Hazelbauer, "Transmembrane signaling in bacterial chemoreceptors.," Trends in biochemical sciences, vol. 26, no. 4, pp. 257–65, 2001.

[107] S. Kitanovic, P. Ames, and J. S. Parkinson, "Mutational analysis of the control cable that mediates transmembrane signaling in the *Escherichia coli* serine chemoreceptor," Journal of bacteriology, vol. 193, no. 19, pp. 5062–5072, 2011.

[108] G. A. Wright, R. L. Crowder, R. R. Draheim, and M. D. Manson, "Mutational analysis of the transmembrane helix 2-HAMP domain connection in the *Escherichia coli* aspartate chemoreceptor Tar," Journal of bacteriology, vol. 193, no. 1, pp. 82–90, 2011.

[109] S. Kitanovic, P. Ames, and J. S. Parkinson, "A trigger residue for transmembrane signaling in the *Escherichia coli* serine chemoreceptor," Journal of bacteriology, vol. 197, no. 15, pp. 2568–2579, 2015.

[110] J. S. Parkinson, "Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases.," Annual review of microbiology, vol. 64, pp. 101–22, 2010.

[111] M. Hulko, F. Berndt, M. Gruber, J. U. Linder, V. Truffault, A. Schultz, J. Martin, J. E. Schultz, A. N. Lupas, and M. Coles, "The HAMP domain structure implies helix rotation in transmembrane signaling," Cell, vol. 126, no. 5, pp. 929–940, 2006.

[112] M. V. Airola, N. Sukomon, D. Samanta, P. P. Borbat, J. H. Freed, K. J. Watts, and B. R. Crane, "HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors.," PLoS biology, vol. 11, no. 2, p. e1001479, 2013.

[113] H. U. Ferris, S. Dunin-Horkawicz, L. G. Mondéjar, M. Hulko, K. Hantke, J. Martin, J. E. Schultz, K. Zeth, A. N. Lupas, and M. Coles, "The mechanisms of HAMP-mediated signaling in transmembrane receptors," Structure, vol. 19, no. 3, pp. 378–385, 2011.

[114] Q. Zhou, P. Ames, and J. S. Parkinson, "Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors.," Molecular microbiology, vol. 73, no. 5, pp. 801–14, 2009.

[115] J. H. Brown, C. Cohen, and D. a. Parry, "Heptad breaks in alpha-helical coiled coils: stutters and stammers.," Proteins, vol. 26, no. 2, pp. 134–45, 1996.

[116] M. D. Coleman, R. B. Bass, R. S. Mehan, and J. J. Falke, "Conserved glycine residues in the cytoplasmic domain of the aspartate receptor play essential roles in kinase coupling and on-off switching.," Biochemistry, vol. 44, no. 21, pp. 7687–95, 2005.

[117] P. Ames, C. a. Studdert, R. H. Reiser, and J. S. Parkinson, "Collaborative signaling by mixed chemoreceptor teams in *Escherichia coli*.," Proceedings of the National Academy of Sciences, vol. 99, no. 10, pp. 7060–5, 2002.

[118] N. L. Bartelli and G. L. Hazelbauer, "Direct evidence that the carboxyl-terminal sequence of a bacterial chemoreceptor is an unstructured linker and enzyme tether.," Protein science: A publication of the protein society, vol. 20, no. 11, pp. 1856–66, 2011.

[119] Q. Zhou, P. Ames, and J. S. Parkinson, "Biphasic control logic of HAMP domain signalling in the *Escherichia coli* serine chemoreceptor," Molecular microbiology, vol. 80, no. 3, pp. 596–611, 2011.

[120] K. E. Swain, M. a. Gonzalez, and J. J. Falke, "Engineered socket study of signaling through a four-helix bundle: evidence for a yin-yang mechanism in the kinase control module of the aspartate receptor.," Biochemistry, vol. 48, no. 39, pp. 9266–77, 2009.

[121] D. R. Ortega, C. Yang, P. Ames, J. Baudry, J. S. Parkinson, and I. B. Zhulin, "A phenylalanine rotameric switch for signal-state control in bacterial chemoreceptors.," Nature communications, vol. 4, p. 2881, 2013.

[122] H. Tajima, K. Imada, M. Sakuma, F. Hattori, T. Nara, N. Kamo, M. Homma, and I. Kawagishi, "Ligand specificity determined by differentially arranged common ligand-binding residues in bacterial amino acid chemoreceptors Tsr and Tar.," The Journal of biological chemistry, vol. 286, no. 49, pp. 42200–10, 2011.

[123] J. U. Bowie, A. A. Pakula, and M. I. Simon, "The three-dimensional structure of the aspartate receptor from *Escherichia coli*," Acta Crystallographica-Section D Biological Crystallography, vol. 51, no. 2, pp. 145–154, 1995.

[124] H. U. Ferris, K. Zeth, M. Hulko, S. Dunin-Horkawicz, and A. N. Lupas, "Axial helix rotation as a mechanism for signal regulation inferred from the crystallographic analysis of the *E. coli* serine chemoreceptor," Journal of structural biology, vol. 186, no. 3, pp. 349–356, 2014.

[125] M. V. Milburn, G. G. Privé, D. L. Milligan, W. G. Scott, J. Yeh, J. Jancarik, D. E. Koshland, and S. H. Kim, "Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand.," Science, vol. 254, no. 5036, pp. 1342–7, 1991.

[126] S. A. Chervitz and J. J. Falke, "Molecular mechanism of transmembrane signaling by the aspartate receptor: a model," Proceedings of the National Academy of Sciences, vol. 93, no. 6, pp. 2545–2550, 1996.

[127] M. L. Peach, G. L. Hazelbauer, and T. P. Lybrand, "Modeling the transmembrane domain of bacterial chemoreceptors.," Protein science: A publication of the protein society, vol. 11, pp. 912–923, 2002.

[128] H. Park, W. Im, and C. Seok, "Transmembrane signaling of chemotaxis receptor Tar: insights from molecular dynamics simulation studies.," Biophysical journal, vol. 100, no. 12, pp. 2955–63, 2011.

[129] B. A. Hall, J. P. Armitage, and M. S. P. Sansom, "Transmembrane helix dynamics of bacterial chemoreceptors supports a piston model of signalling.," PLoS computational biology, vol. 7, no. 10, p. e1002204, 2011.

[130] A. Pakula and M. Simon, "Determination of transmembrane protein structure by disulfide cross-linking: The *Escherichia coli* Tar receptor," Proceedings of the National Academy of Sciences, vol. 89, no. May, pp. 4144–4148, 1992.

[131] A. Drozdetskiy, C. Cole, J. Procter, and G. J. Barton, "Jpred4: A protein secondary structure prediction server," Nucleic acids research, p. gkv332, 2015.

[132] X. Chen and D. E. Koshland, "The N-terminal cytoplasmic tail of the aspartate receptor is not essential in signal transduction of bacterial chemotaxis," Journal of Biological Chemistry, vol. 270, no. 41, pp. 24038–24042, 1995.

[133] B. Webb and A. Sali, "Comparative protein structure modeling using Modeller," Current protocols in bioinformatics, pp. 5–6, 2014.

[134] E. Wu, X. Cheng, S. Jo, H. Rui, K. Song, E. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. Venable, J. Klauda, and W. Im, "CHARMM-GUI Membrane Builder toward realistic biological membrane simulations," Journal of computational chemistry, vol. 35, no. 27, pp. 1997–2004, 2014.

[135] D. N. Amin and G. L. Hazelbauer, "Influence of membrane lipid composition on a transmembrane bacterial chemoreceptor.," The Journal of biological chemistry, vol. 287, no. 50, pp. 41697–705, 2012.

[136] S. R. Eddy, "Where did the BLOSUM62 alignment score matrix come from?," Nature biotechnology, vol. 22, no. 8, pp. 1035–1036, 2004.

[137] A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton, "Jalview version 2, a multiple sequence alignment editor and analysis workbench," Bioinformatics, vol. 25, no. 9, pp. 1189–1191, 2009.

[138] J. J. Falke and D. E. Koshland, "Global flexibility in a sensory receptor: a site-directed cross-linking approach," Science, vol. 237, no. 4822, pp. 1596–1600, 1987.

[139] S. A. Chervitz and J. J. Falke, "Lock on/off disulfides identify the transmembrane signaling helix of the aspartate receptor," Journal of Biological Chemistry, vol. 270, no. 41, pp. 24043–24053, 1995.

[140] A. S. Miller and J. J. Falke, "Side chains at the membrane-water interface modulate the signaling state of a transmembrane receptor," Biochemistry, vol. 43, no. 7, pp. 1763–1770, 2004.

[141] R. R. Draheim, A. F. Bormans, R.-Z. Lai, and M. D. Manson, "Tuning a bacterial chemoreceptor with protein-membrane interactions," Biochemistry, vol. 45, no. 49, pp. 14655–14664, 2006.

[142] I. T. Arkin and A. T. Brunger, "Statistical analysis of predicted transmembrane α-helices," Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, vol. 1429, no. 1, pp. 113–128, 1998.

[143] E. Wallin, T. Tsukihara, S. Yoshikawa, G. V. Heijne, and A. Elofsson, "Architecture of helix bundle membrane proteins: an analysis of cytochrome C oxidase from bovine mitochondria," Protein Science, vol. 6, no. 4, pp. 808–815, 1997.

[144] J. A. Killian and G. von Heijne, "How proteins adapt to a membrane–water interface," Trends in biochemical sciences, vol. 25, no. 9, pp. 429–434, 2000.

[145] E. Strandberg and J. A. Killian, "Snorkeling of lysine side chains in transmembrane helices: how easy can it get?," Febs Letters, vol. 544, no. 1-3, pp. 69–73, 2003.

[146] A. C. E. Dahl, M. Chavent, and M. S. Sansom, "Bendix: intuitive helix geometry analysis and abstraction," Bioinformatics, vol. 28, no. 16, pp. 2193–2194, 2012.

[147] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," Nature chemical biology, vol. 5, no. 11, pp. 789–796, 2009.

[148] J. M. Mason and K. M. Arndt, "Coiled coil domains: stability, specificity, and biological implications," ChemBioChem, vol. 5, no. 2, pp. 170–176, 2004.

[149] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles," Journal of chemical theory and computation, vol. 8, no. 9, pp. 3257–3273, 2012.

[150] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr, and R. W. Pastor, "Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types," The journal of physical chemistry B, vol. 114, no. 23, pp. 7830–7843, 2010.

[151] K. Volz, "Structural conservation in the CheY superfamily," Biochemistry, vol. 32, no. 44, pp. 11741–11753, 1993.

[152] H. S. Cho, S.-Y. Lee, D. Yan, X. Pan, J. S. Parkinson, S. Kustu, D. E. Wemmer, and J. G. Pelton, "NMR structure of activated CheY," Journal of molecular biology, vol. 297, no. 3, pp. 543–551, 2000.

[153] K. Volz and P. Matsumura, "Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution.," Journal of Biological Chemistry, vol. 266, no. 23, pp. 15511–15519, 1991.

[154] M. M. McEvoy, A. Bren, M. Eisenbach, and F. W. Dahlquist, "Identification of the binding interfaces on CheY for two of its targets the phosphatase CheZ and the flagellar switch protein FliM," Journal of molecular biology, vol. 289, no. 5, pp. 1423–1433, 1999.

[155] A. Bren and M. Eisenbach, "The N-terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY," Journal of molecular biology, vol. 278, no. 3, pp. 507–514, 1998.

[156] Y. Blat and M. Eisenbach, "Phosphorylation-dependent binding of the chemotaxis signal molecule CheY to its phosphatase, CheZ," Biochemistry, vol. 33, no. 4, pp. 902–906, 1994.

[157] R. Ramakrishnan, M. Schuster, and R. B. Bourret, "Acetylation at Lys-92 enhances signaling by the chemotaxis response regulator protein CheY," Proceedings of the National Academy of Sciences, vol. 95, no. 9, pp. 4918–4923, 1998.

[158] R. Barak, K. Prasad, A. Shainskaya, A. J. Wolfe, and M. Eisenbach, "Acetylation of the chemotaxis response regulator CheY by acetyl-CoA synthetase purified from *Escherichia coli*," Journal of molecular biology, vol. 342, no. 2, pp. 383–401, 2004.

[159] R. Barak, J. Yan, A. Shainskaya, and M. Eisenbach, "The chemotaxis response regulator CheY can catalyze its own acetylation," Journal of molecular biology, vol. 359, no. 2, pp. 251–265, 2006.

[160] R. Li, J. Gu, Y.-Y. Chen, C.-L. Xiao, L.-W. Wang, Z.-P. Zhang, L.-J. Bi, H.-P. Wei, X.-D. Wang, and J.-Y. Deng, "CobB regulates *Escherichia coli* chemotaxis by deacetylating the response regulator CheY," Molecular microbiology, vol. 76, no. 5, pp. 1162–1174, 2010.

[161] S. Thao and J. C. Escalante-Semerena, "Control of protein function by reversible N-lysine acetylation in bacteria," Current opinion in microbiology, vol. 14, no. 2, pp. 200–204, 2011.

[162] G.-W. Kim and X.-J. Yang, "Comprehensive lysine acetylomes emerging from bacteria to humans," Trends in biochemical sciences, vol. 36, no. 4, pp. 211–220, 2011.

[163] L. Ma and Q. Cui, "Activation mechanism of a signaling protein at atomic resolution from advanced computations," Journal of the American Chemical Society, vol. 129, no. 33, pp. 10261–10268, 2007.

[164] J. M. Mottonen, D. J. Jacobs, and D. R. Livesay, "Allosteric response is both conserved and variable across three CheY orthologs," Biophysical journal, vol. 99, no. 7, pp. 2245–2254, 2010.

[165] M. H. Knaggs, F. R. Salsbury, M. H. Edgell, and J. S. Fetrow, "Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations," Biophysical journal, vol. 92, no. 6, pp. 2062–2079, 2007.

[166] A. Wolf and K. N. Kirschner, "Principal component and clustering analysis on molecular dynamics data of the ribosomal l11 and 23s subdomain," Journal of molecular modeling, vol. 19, no. 2, pp. 539–549, 2013.

[167] L. Skjaerven, A. Martinez, and N. Reuter, "Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit," Proteins: Structure, Function, and Bioinformatics, vol. 79, no. 1, pp. 232–243, 2011.

[168] C. Ding and X. He, "K-means clustering via principal component analysis," in Proceedings of the Twenty-first International Conference on Machine Learning (C. Brodley, ed.), pp. 29–38, New York: ACM, 2004.

[169] A. Bakan, L. M. Meireles, and I. Bahar, "ProDy: protein dynamics inferred from theory and experiments," Bioinformatics, vol. 27, no. 11, pp. 1575–1577, 2011.